

Galaxy and Mass Assembly (GAMA): the clustering of galaxy groups

Article (Published Version)

Riggs, S D, Barbhuiyan, R W Y M, Loveday, J, Brough, S, Holwerda, B W, Hopkins, A M and Phillipps, S (2021) Galaxy and Mass Assembly (GAMA): the clustering of galaxy groups. Monthly Notices of the Royal Astronomical Society, 506 (1). pp. 21-37. ISSN 0035-8711

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/100816/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Galaxy and mass assembly (GAMA): the clustering of galaxy groups

S. D. Riggs^{1b},^{1★} R. W. Y. M. Barbhuiyan,¹ J. Loveday^{1b},¹ S. Brough^{1b},² B. W. Holwerda^{1b},³
A. M. Hopkins⁴ and S. Phillipps^{1b5}

¹*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK*

²*School of Physics, University of New South Wales, NSW 2052, Australia*

³*Department of Physics and Astronomy, University of Louisville, Louisville, KY 40292, USA*

⁴*Australian Astronomical Optics, Macquarie University, 105 Delhi Road, North Ryde, NSW 2113, Australia*

⁵*Astrophysics Group, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK*

Accepted 2021 June 10. Received 2021 June 10; in original form 2021 February 8

ABSTRACT

We explore the clustering of galaxy groups in the Galaxy and Mass Assembly (GAMA) survey to investigate the dependence of group bias and profile on separation scale and group mass. Due to the inherent uncertainty in estimating the group selection function, and hence the group autocorrelation function, we instead measure the projected galaxy–group cross-correlation function. We find that the group profile has a strong dependence on scale and group mass on scales $r_{\perp} \lesssim 1 h^{-1}$. We also find evidence that the most massive groups live in extended, overdense, structures. In the first application of marked clustering statistics to groups, we find that group-mass marked clustering peaks on scales comparable to the typical group radius of $r_{\perp} \approx 0.5 h^{-1}$. While massive galaxies are associated with massive groups, the marked statistics show no indication of galaxy mass segregation within groups. We show similar results from the IllustrisTNG simulations and the L-GALAXIES model, although L-GALAXIES shows an enhanced bias and galaxy mass dependence on small scales.

Key words: galaxies: groups: general – galaxies: haloes – large-scale structure of Universe.

1 INTRODUCTION

In the standard hierarchical model of galaxy formation, galaxies form in gravitationally collapsed dark matter (DM) haloes which grow by merging with other haloes (e.g. Press & Schechter 1974; White & Rees 1978). Consequently, the relative density of observable matter (δ_g , such as galaxies, galaxy groups, and galaxy clusters) in a given volume of space is believed to trace the relative density of dark matter, δ_m , in that same space. In the linear bias model, $\delta_g = b\delta_m$, where b is known as the *bias* parameter, which will in general be a function of the tracer population, separation scale, and redshift. This linear bias has previously been shown to increase with halo mass (e.g. Mo & White 1996; Sheth & Tormen 1999; Sheth, Mo & Tormen 2001; Seljak & Warren 2004; Tinker et al. 2005).

A direct way to explore the connection between galaxies and their DM haloes is with galaxy group catalogues. The total mass of individual haloes can be estimated using the galaxy motions within them (e.g. Girardi et al. 1998; Eke et al. 2006; Robotham et al. 2011), or by scaling relations based on the luminosity or mass of their constituent galaxies (e.g. Yang et al. 2007; Han et al. 2015; Viola et al. 2015). The galaxy distribution within haloes can be explored directly by group stacking (e.g. Budzynski et al. 2012) or with group-galaxy clustering (e.g. Wang et al. 2008; Mohammad et al. 2016). Group clustering probes intermediate scales compared to the typical galaxy- and galaxy cluster-scales used in most clustering studies, and

can be combined with galaxy- and dark matter-clustering to extract estimates of bias.

The mass and colour dependence of the clustering and bias of galaxy groups was investigated for *Sloan Digital Sky Survey* (SDSS) Data Release 4 (Adelman-McCarthy et al. 2006) by Wang et al. (2008). They found that the clustering strength of groups increases with increasing total group mass and also that groups of comparable mass are more strongly clustered when they contain redder galaxies. Similar results from SDSS were found in the earlier study by Berlind et al. (2006), where a sharp increase in group-galaxy clustering is observed within the typical group scale compared to larger scales. Further, group-galaxy clustering is observed to decrease slightly on scales $r_{\perp} \lesssim 0.3 h^{-1}$ Mpc, possibly suggesting the existence of group cores, although clustering measurements on these scales are sensitive to the choice of group centre, a point we discuss further in Section 5.1. An increase in clustering strength with increasing group mass has also been shown at slightly higher redshifts using the zCOSMOS survey (Lilly et al. 2007) by Knobel et al. (2012).

While there is a general consensus in previous work on the group clustering increase with group mass at large scale, the details on small scales are less constrained. A key aspect to this is the dependence of the positions of galaxies within groups on the properties of the satellite galaxies. Mass segregation, a tendency for more massive galaxies to be closer to the group centre, is found by, e.g. Presotto et al. (2012), Roberts et al. (2015), but other studies (e.g. von der Linden et al. 2010; Kafle et al. 2016) find no trend in stellar mass with radial distance from group centre. The presence or absence of mass segregation helps constrain the strength of dynamical friction

★ E-mail: S.Riggs@sussex.ac.uk

effects within haloes, as satellites infall at large radii (Wetzel et al. 2013) and then move inwards due to dynamical friction.

Standard two-point clustering measurements can be expanded on using marked statistics (Stoyan & Stoyan 1994; Beisbart & Kerscher 2000; Sheth & Tormen 2004; Sheth, Connolly & Skibba 2005; Harker et al. 2006; Skibba et al. 2006; White & Padmanabhan 2009; White 2016). These have been used to explore the environmental dependence of clustering, with Skibba et al. (2013) finding that small-scale clustering is dependent on local density, and Sheth & Tormen (2004) showing that close pairs of haloes form earlier. Armijo et al. (2018) show that galaxy clustering has an increasing dependence on halo mass on smaller scales. However, this method has not to our knowledge previously been applied to the exploration of group clustering.

The Galaxy and Mass Assembly (GAMA; Driver et al. 2009, 2011; Liske et al. 2015) survey provides an opportunity to reassess the clustering of galaxy groups. GAMA has a smaller area than SDSS, but provides spectroscopic redshifts two magnitudes fainter (Hopkins et al. 2013), and is highly complete, even in the high-density environments of galaxy groups. We thus expect the GAMA group catalogue to be more reliable than group catalogues constructed from SDSS data, and to allow the exploration of group clustering on much smaller scales. The clustering of GAMA galaxies has been shown to increase with luminosity and mass by Farrow et al. (2015). The dependency of galaxy clustering in GAMA on galaxy properties has been explored with marked correlation functions by Gunawardhana et al. (2018) and Sureshkumar et al. (2021), finding that specific star formation rate best traces interactions, and stellar mass best traces environment. Within GAMA groups, Kafle et al. (2016) find negligible mass segregation for satellites. Recently, Vázquez-Mata et al. (2020, hereafter VM20) explored the stellar masses and r -band luminosities of galaxies in GAMA groups, finding brighter and more massive galaxies in more massive groups.

In this paper, we present group–galaxy cross-correlation functions from the GAMA survey; exploring their dependence on scale and group mass. We consider both the large, intergroup, scales which can be compared to results from SDSS, and the smaller, intragroup, scales that are opened up with the high completeness of GAMA. We further examine these dependencies by presenting the first application of marked correlations to group clustering. We also compare these correlation functions to results from the IllustrisTNG hydrodynamical simulations (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018, 2019; Pillepich et al. 2018; Springel et al. 2018) and the L-GALAXIES semi-analytic model (Henriques et al. 2015).

The layout of this paper is as follows: in Section 2, we describe the data selection from the GAMA survey, mock catalogues and models we compare against; in Section 3, we detail the methods used to derive the two-point correlation functions and marked statistics; in Section 4, we present our results; and finally in Sections 5 and 6 we provide a discussion and conclusion. The cosmology assumed throughout is that of a Λ CDM model with $\Omega_\Lambda = 0.75$, $\Omega_m = 0.25$, and $H_0 = h 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. We represent group (halo) masses on a logarithmic scale by $\lg \mathcal{M}_h \equiv \log_{10}(\mathcal{M}_h / \mathcal{M}_\odot h^{-1})$, where we take \mathcal{M}_h to be M_{200} , defined by the mass enclosed within an overdensity 200 times the mean density of the Universe.

2 DATA, MOCKS, AND SIMULATIONS

The GAMA data and mock catalogues used in this analysis are identical to those used in a recent study of the dependence of the galaxy luminosity and stellar mass functions on the mass of their host groups (VM20), although we select groups and galaxies using

different mass and redshift cuts. We summarize the most salient features here.

We make use of the GAMA-II (Liske et al. 2015) equatorial fields G09, G12, and G15, centred on 09^{h} , 12^{h} , and $14^{\text{h}}30^{\text{m}}$ RA, respectively. These fields each have an area of $12 \times 5 \text{ deg}$, Petrosian magnitude limit of $r < 19.8 \text{ mag}$, and a completeness greater than 96 per cent for all galaxies which have up to 5 neighbours within 40 arcsec; for a more in-depth description see Liske et al. (2015).

2.1 Galaxy sample

It is necessary to use a volume-limited sample of galaxies for cross-correlating with groups, as more massive groups are at higher redshift, where galaxies in a flux-limited sample will be more luminous and therefore more strongly clustered. In other words, using a flux-limited galaxy sample, apparent clustering strength would increase with halo mass, even if there was no dependence of halo bias on mass.

We select a volume-limited sample of 42 679 GAMA-II galaxies which have $(K + e)$ -corrected r -band Petrosian magnitude $^{0.1}M_r < -20 \text{ mag}$, with corresponding redshift limit $z_{\text{lim}} < 0.267$ and mean number density $n = 5.38 \times 10^{-3} h^3 \text{ Mpc}^{-3}$. This corresponds to the ‘V0’ sample of Loveday et al. (2018, hereafter L18),¹ and is chosen to roughly maximize survey volume and number of galaxies. We choose to define the volume-limited sample by luminosity rather than stellar mass, as (i) the parent sample is magnitude limited, meaning that variations in mass-to-light ratio would require much more stringent cuts on mass than on luminosity, and (ii) estimated stellar mass is inherently more uncertain (and model dependent) than luminosity.

To account for the different redshifts at which galaxies are observed, the intrinsic luminosities of the GAMA galaxies we use have been corrected by the so-called K -correction (Humason, Mayall & Sandage 1956). We obtain K -corrections from the GAMA data management unit (DMU) `kCorrectionsv05`; see Loveday et al. (2015) for details on how these were calculated. We K -correct to a passband blueshifted by $z = 0.1$ in order to minimize the size, and hence uncertainty, in K -correction. Absolute magnitudes in this band are indicated by $^{0.1}M_r$. We include luminosity evolution by applying correction of $+Q_e z \text{ mag}$, where $Q_e = 1.0$.

The statistics of the GAMA volume-limited galaxy sample, along with those of the mock catalogue and simulations, are summarized in Table 1. The GAMA data, mocks and simulations have galaxies with differing K and e -corrections, and different luminosity functions, but we only need luminosities in order to generate comparable volume-limited galaxy samples. We therefore choose magnitude limits (shown in the second column of Table 1), in order to achieve approximately equal number densities (final column), and hence clustering properties. Note that the GAMA mocks were designed to have a luminosity function very close to that of the GAMA data (R11). The different magnitude limits in Table 1 most likely reflect differences in the K - and e - corrections assumed, as well as sample variance in the GAMA data (Driver et al. 2011). For reference, the clustering and stellar mass distribution for our GAMA, mock, and simulated galaxy samples are presented in Appendix A.

¹ Attentive readers will notice that here we use a slightly higher redshift limit for the same absolute magnitude limit as L18. This is due to an alternative way of defining a 95 percent complete sample. In L18, we take the 95th-percentile of the K -correction of galaxies within $z_{\text{lim}} \pm 0.01$. Here, we take the 95th-percentile of the projected K -correction $K(z_{\text{lim}})$ of all galaxies with $z < z_{\text{lim}}$.

Table 1. Definition of galaxy volume-limited samples for GAMA data, mocks, TNG300-1 simulation, and L-GALAXIES SAM. The columns are absolute r -band magnitude limit (K -corrected to redshift 0.1 for GAMA, redshift 0.0 for other samples), redshift limit, sample volume, number of galaxies selected, and mean density. GAMA data and mocks cover areas of 180 and 144 deg², respectively. The mock sample was volume-limited to redshift 0.301 before applying the GAMA redshift limit, leading to a slightly higher final number density. TNG300-1 and L-GALAXIES use periodic boxes, and so are volume-limited by nature. The redshifts we quote for them are those of the output snapshot used.

	M_{lim}	z_{lim}	V ($10^6 h^{-3} \text{Mpc}^3$)	N_{gal}	\bar{n} ($10^{-3} h^3 \text{Mpc}^{-3}$)
GAMA	−20.00	0.267	7.93	42 679	5.38
Mock	−20.21	0.267	6.35	34 615	5.45
TNG300-1	−19.83	0.200	8.62	46 349	5.38
L-GALAXIES	−20.12	0.180	110.78	596 023	5.38

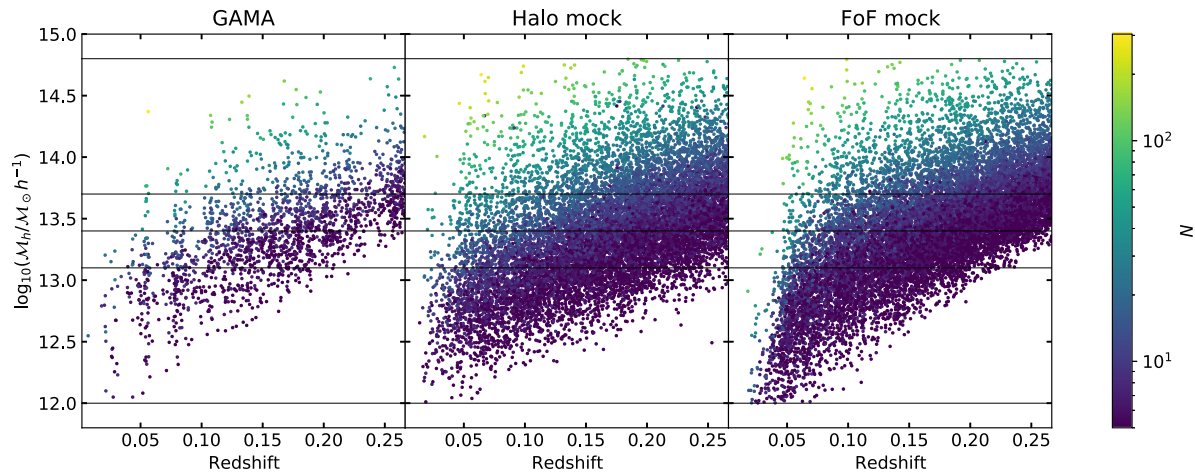


Figure 1. Mass–redshift distribution for GAMA and mock groups at $z < 0.267$ with at least 5 members. Groups are colour coded by the number of group members. The horizontal lines show the division of groups into halo mass bins. Mock groups are shown for all nine realizations of the lightcone.

2.2 GAMA groups

The GAMA Galaxy Group Catalogue (G³Cv9) was produced by grouping galaxies in the GAMA-II spectroscopic survey using a friends-of-friends (FoF) algorithm; this is an updated version of G³Cv1 which was generated from the GAMA-I survey by Robotham et al. (2011, hereafter R11). The FoF parameters used for G³Cv9 (hereafter abbreviated to G³C) are identical to those in R11, but applied to the larger GAMA-II galaxy sample. G³C contains 23 654 groups with 2 or more members and overall ~ 40 per cent of galaxies in GAMA are assigned to G³C groups. In this study, we utilize only those groups within the redshift limit $z < 0.267$ of our volume-limited galaxy sample (Table 1), and which have five or more member galaxies, as R11 find these richer groups are most reliable. Reducing the threshold on the number of group members increases the number of low-mass groups, but these groups are very unreliable as chance alignments are increasingly included in the group sample. We also require groups in our sample to have $\text{GroupEdge} > 0.9$, this removes any where it is estimated that less than 90 per cent of the group is within the GAMA-II survey boundaries. This leaves us with a sample of 1894 groups with $12.0 < \lg M_h < 14.8$. We do not attempt to form a volume-limited sample of GAMA groups, as selection effects are complex (see VM20 for a discussion), and to do so would severely limit the number of groups that could be used.

We take the centre of these groups to be the iterative central from R11, found by iteratively removing galaxies from the centre of light until one is left. We choose this as it is found by R11 to be the best

estimator of true central, but we discuss the choice of this further in Section 5.1.

Halo masses M_h are estimated from group r -band luminosity (column LumB) using the scaling relation for M_{200} of Viola et al. (2015, equation 37), which was calibrated against weak-lensing measurements. The LumB column contains total r -band luminosities down to $M_r - 5\log_{10}h = -14$ mag in solar luminosities, corrected by an empirical factor B which has been calibrated against mock catalogues (see R11 section 4.4 for details). The G³C also provides dynamical mass estimates derived via the virial theorem (column MassA).

Our choice of luminosity-based mass estimates follows the checks on mass estimate reliability by VM20, who find that the luminosity-based estimates correlate much better with true halo mass than dynamical mass estimates (VM20 Fig. 1).

We show the mass–redshift distribution of our selected GAMA groups in the left-hand panel of Fig. 1. Due to the $r < 19.8$ mag flux limit of GAMA-II and our requirement for groups to contain at least 5 members, low-mass groups are less likely to be detected at higher redshifts, and the groups that are detected generally have fewer observed members.

We sub-divide the groups into four mass bins as defined in Table 2, chosen as a compromise between bins of fixed mass range and comparable group numbers. As seen in VM20, the central galaxy luminosity is greater for more massive groups, with our mass bins M1–4 having central galaxy mean absolute magnitudes of $^{0.1}M_r - 5\log_{10}h = -20.48, -21.12, -21.48, \text{ and } -21.87$, respectively.

Table 2. Group bin names and log-mass limits, number of groups, mean log-mass, and mean redshift for GAMA-II data, intrinsic mock haloes, FoF mock groups, TNG300-1 simulation haloes, and L-GALAXIES SAM haloes. Note that the values given for mocks are averages across the 9 realizations used. TNG300-1 and L-GALAXIES are results from single snapshots down sampled to select groups comparable to GAMA.

	$\lg \mathcal{M}_{h, \text{limits}}$	N	GAMA		N	Halo mocks		N	FoF mocks		N	TNG300-1		N	L-GALAXIES	
			$\lg \mathcal{M}$	\bar{z}		$\lg \mathcal{M}$	\bar{z}		$\lg \mathcal{M}$	\bar{z}		$\lg \mathcal{M}$	\bar{z}		$\lg \mathcal{M}$	\bar{z}
$\mathcal{M}1$	[12.0, 13.1]	380	12.87	0.10	352	12.86	0.11	346	12.80	0.10	414	12.84	0.10	5276	12.84	0.10
$\mathcal{M}2$	[13.1, 13.4]	547	13.26	0.15	383	13.25	0.16	401	13.26	0.15	383	13.25	0.15	4986	13.25	0.15
$\mathcal{M}3$	[13.4, 13.7]	566	13.54	0.19	366	13.54	0.19	523	13.55	0.19	405	13.54	0.19	5127	13.55	0.19
$\mathcal{M}4$	[13.7, 14.8]	401	13.93	0.20	306	13.97	0.20	430	13.96	0.21	461	13.98	0.21	6263	14.00	0.21
Total	[12.0, 14.8]	1894	13.41	0.16	1407	13.39	0.17	1699	13.43	0.17	1663	13.42	0.17	21652	13.44	0.17

We note that this means that the $\mathcal{M}1$ centrals have a lower mean luminosity than our volume limited galaxy sample, which has a mean $^{0.1}M_r - 5\log_{10}h = -20.59$, and so the $\mathcal{M}1$ groups are expected to be slightly less clustered than the galaxy sample.

2.3 Mock catalogues

We compare our results with predictions from two sets of mock group catalogues for the GAMA-I survey (catalogues updated to the GAMA-II survey are currently being developed). These catalogues were produced using lightcones from the GALFORM (Bower et al. 2006) semi-analytic galaxy formation model (SAM) run on the Millennium dark matter simulation (Springel et al. 2005). For more details on these mocks we refer the reader to R11.

The first set of mocks are G3CMockHaloGroupv06, which we refer to as *halo mocks*. This contains the dark matter haloes in the simulations, with their positions and masses $\mathcal{M}_{\text{Dhalo}}$. The definition of $\mathcal{M}_{\text{Dhalo}}$ differs slightly from M_{200} , but Jiang et al. (2014) and R11 find they are median unbiased relative to each other, so we can use $\mathcal{M}_{\text{Dhalo}}$ as an estimate of M_{200} . The second set of mocks are G3CMockFoFGroupv06, which we refer to as *FoF mocks*. The groups in this are generated with the same FoF algorithm as GAMA, and masses \mathcal{M}_{lum} estimated using the same Viola et al. (2015) luminosity scaling relation. Comparing results from these two mock group catalogues thus allows us to assess the impact on estimated halo clustering of redshift-space group-finding and luminosity-based mass estimation. For halo and FoF mock groups that share a common central galaxy, Fig. 2 compares \mathcal{M}_{lum} with $\mathcal{M}_{\text{Dhalo}}$. The upper panel shows all groups with $N_{\text{FoF}} \geq 5$ (and implicitly $N_{\text{halo}} \geq 2$ to be counted as a group), while the lower panel shows groups with $N_{\text{FoF}} \geq 5$ and $N_{\text{halo}} \geq 5$. From the lower panel, it is apparent there is reasonable agreement of \mathcal{M}_{lum} to $\mathcal{M}_{\text{Dhalo}}$ (within one standard deviation) for groups that have sufficient members to be included in our halo catalogue sample. However, it is clear from the upper panel that there is a population of groups which have their membership, and therefore mass, overestimated in the FoF mocks. Through the rest of this work, for consistency with our GAMA selection, we use all groups in the FoF mock with $N_{\text{FoF}} \geq 5$, so the groups with overestimated mass are included. In the halo mock we select all groups with $N_{\text{halo}} \geq 5$, representing the sample we would have if the FoF group finder perfectly assigned galaxies to groups.

The central and right-hand panels of Fig. 1, showing the mass-redshift relation for all selected mock groups, further shows that the luminosity-based masses have a stronger redshift dependence than the true halo masses. The mass overestimation appears to be greater at high redshift. However, at redshifts $z \lesssim 0.04$ the FoF mock groups mostly have low masses, suggesting galaxies are missed from the outskirts of the more extended groups at low redshift. We expect this

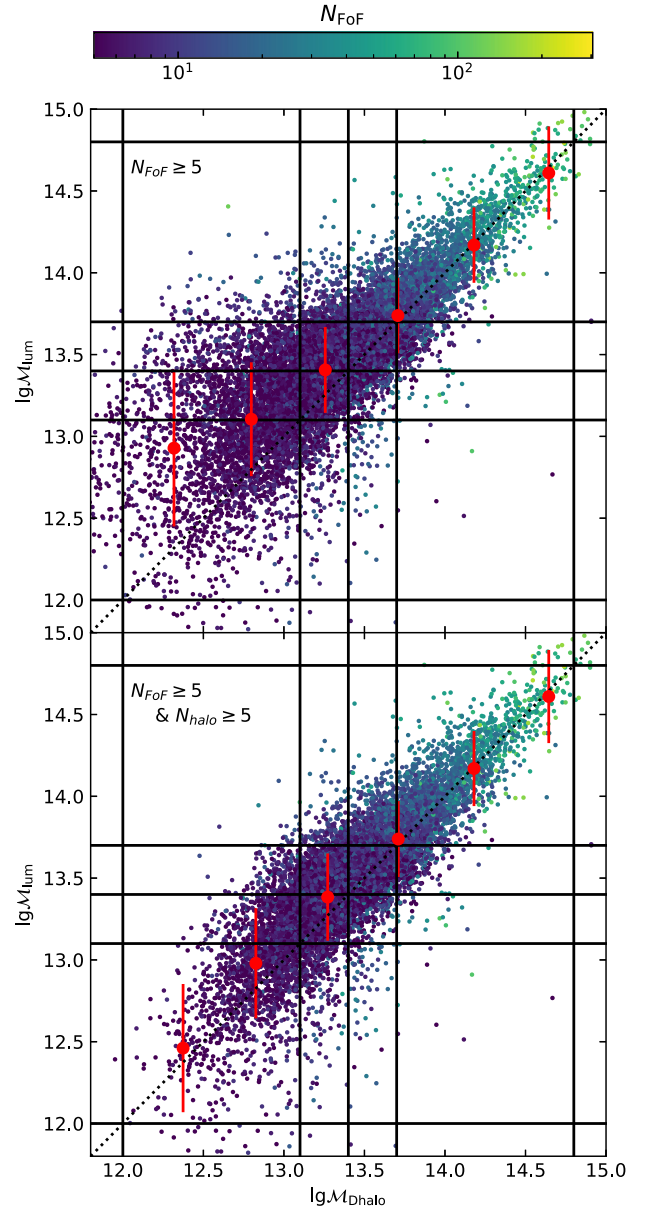


Figure 2. Comparison of luminosity-based ($\lg \mathcal{M}_{\text{lum}}$) estimates of mock group mass, against true mock halo mass ($\lg \mathcal{M}_{\text{Dhalo}}$), colour coded by group membership, for groups at redshifts $z < 0.267$. The upper panels show groups selected by their visibility in the FoF mocks ($N_{\text{FoF}} \geq 5$), while the lower panels show only those groups visible in both mocks ($N_{\text{FoF}} \geq 5$ and $N_{\text{halo}} \geq 5$). The red errorbars show mean and standard deviation of $\lg \mathcal{M}_{\text{lum}}$ in 0.5 mag bins of $\lg \mathcal{M}_{\text{Dhalo}}$. The horizontal and vertical lines delineate the halo mass bins used in this analysis.

to imply a similar trend in group mass misestimation with redshift will also be present in the groups from GAMA.

Mock galaxies are taken from the galaxy catalogue associated with the mock groups we use, G3CMockGalv06. We K -correct the absolute magnitudes to redshift zero with the K - and e -corrections specified in section 2.2 of R11. Due to differences with GAMA K - and e -corrections, we set the galaxy magnitude limit by trial-and-error to give approximately the same mean volume-limited number density as the GAMA galaxy sample. This results in a sample with a limiting absolute magnitude $^{0.0}M_r < -20.21$ and limiting redshift $z_{\text{lim}} < 0.301$. The typical masses of observed galaxies and groups increase with redshift, and so to ensure that the mock samples are comparable to the observations, we then restrict our mock sample to the GAMA redshift limit of $z_{\text{lim}} < 0.267$. The details of our final mock galaxy sample are given in Table 1.

We estimate uncertainties on mock clustering from the scatter between nine realizations of the GAMA-I survey equatorial regions. Each of these realizations consists of three 12×4 deg regions; which are 20 per cent smaller in area (and so also volume) than the equatorial fields we use from GAMA-II. Galaxy stellar masses are not included in these mocks so we cannot explore the dependence of the marked correlation on galaxy mass in the mocks.

2.4 Random catalogues

A random sample of points is needed to model any selection effects in the galaxy sample (our choice of cross-correlation estimator in Section 3.1 means that the selection function of group samples is not needed). We use the same survey mask described in section 2.3.1 of L18, and generate angular coordinates using MANGLE (Hamilton & Tegmark 2004; Swanson et al. 2008). Radial coordinates are drawn at random from a uniform distribution in comoving volume with a modulation factor of $10^{0.4Pz}$, the density-evolution factor of Loveday et al. (2015, equation 5), taking $P = 1$. We generate 10 times more random points than galaxies.

In Fig. 3, we show galaxy redshift distributions for GAMA, the average across the nine mocks, and random samples (with the number of randoms divided by 10 to match the data samples). The random number counts accurately reproduce the GAMA redshift distribution except for fluctuations due to large-scale structure (cf. Loveday et al. 2015, Fig. 7).

2.5 Comparison models

In addition to comparison with GAMA mock catalogues, we also compare our results with predictions from the IllustrisTNG hydrodynamical simulations (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018, 2019; Pillepich et al. 2018; Springel et al. 2018) and the Henriques et al. (2015) version of the L-GALAXIES semi-analytic model. For each of these, we select galaxies at a snapshot close to the GAMA mean redshift, selecting $z = 0.20$ in IllustrisTNG and $z = 0.18$ (the closest snapshot to $z = 0.20$) in L-GALAXIES, and set the absolute magnitude limit of the galaxy sample in order to give the same approximate number density as the GAMA volume-limited sample, viz., $n = 5.38 \times 10^{-3} h^3 \text{Mpc}^{-3}$.

For IllustrisTNG, we use the highest resolution simulation at the largest box-size of 300 Mpc ($205 h^{-1} \text{Mpc}$ for $h = 0.6774$), TNG300-1. Haloes are selected by M_{200} (Group.M.Mean200) using the mass limits in Table 2. For galaxy masses we select the stellar component (type 4) of the SubhaloMassInRadType field, which gives the stellar mass within twice the stellar half-mass radius. Following the recommendation of Pillepich et al. (2018), we multiply these by a

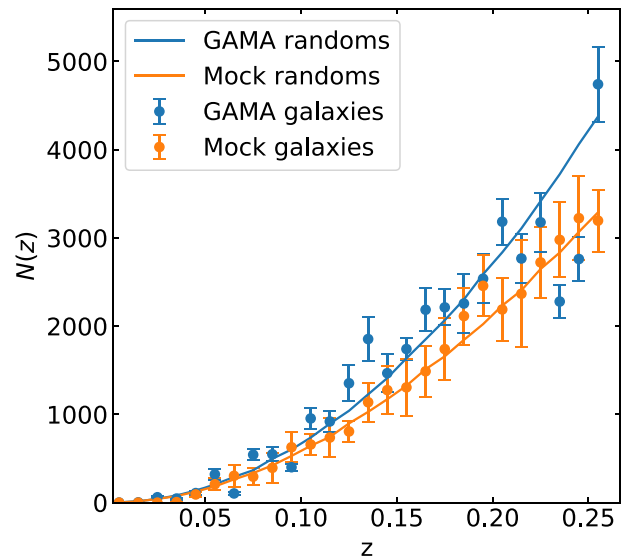


Figure 3. Comparison of galaxy redshift distributions for GAMA, the average across the nine mocks, and the random samples. Random counts have been divided by 10 to account for the larger number of random points generated. Uncertainties on GAMA and random counts are found by jackknife between 27 regions in RA, and on the mock counts by the scatter between 9 realizations. The offset in the number of galaxies between GAMA and the mocks is due to the larger area of GAMA (180 deg^2 compared to 144 deg^2).

factor of 1.4, appropriate for haloes in the mass range $12 < \lg M_h < 15$. We use the dust-corrected luminosities derived from dust model C of Nelson et al. (2018) when selecting the volume-limited galaxy sample.

For L-GALAXIES, we use the Henriques et al. (2015) version with the Millennium (Springel et al. 2005) N -body simulation. Haloes are again selected by M_{200} and the total stellar mass of the galaxies is taken.

To avoid including galaxies below the resolution limits of the TNG300-1 and L-GALAXIES simulations, we select only galaxies with $\log_{10} M_\star > 9.0 M_\odot$.

To provide comparable group samples, we need to allow for the fact that the periodic-cube (i.e. volume-limited) simulations contain many more low-mass groups than the flux-limited GAMA data and mocks. We describe here our approach to the group selection; in Appendix B, we validate our method and demonstrate the consequences of not applying it.

Since we are measuring only group-galaxy cross-correlation functions, we do not require the simulated groups to have an accurate group autocorrelation. Therefore, rather than attempt to create light-cones from the simulations, we simply down-sample the simulated groups to match the mass distribution of selected GAMA groups. We do this by estimating the probability of finding each halo² within the GAMA volume. In our GAMA sample we have set $N_{\text{FoF}} \geq 5$, and so the halo selection probability is dependent on the fifth brightest galaxy in the halo. To calculate this probability and select simulated groups we use the following procedure for each halo:

- (i) Identify the absolute magnitude of the fifth brightest galaxy in the halo.

²The terms ‘halo’ and ‘group’ are used interchangeably when discussing the TNG and L-GALAXIES simulations.

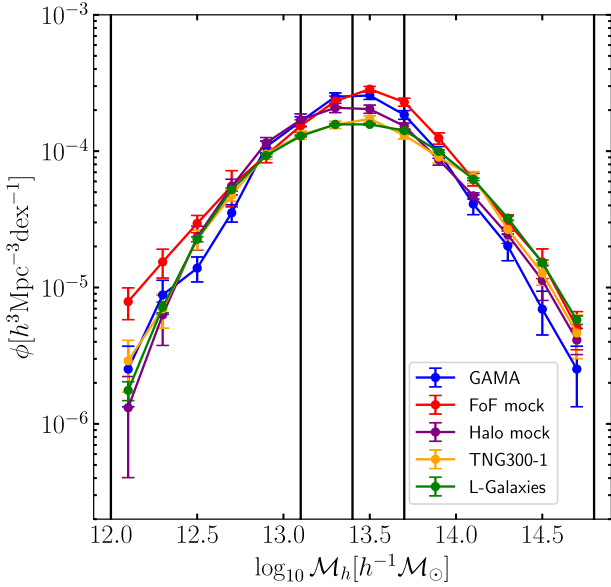


Figure 4. Distribution of group (halo) masses in our sample for GAMA, the two mock catalogues, TNG300-1 and L-GALAXIES. The plotted uncertainties are jackknife values between 27 regions for GAMA and simulations, and the scatter between 9 realizations for the mocks. The vertical lines delineate the halo mass bins used in this analysis.

(ii) Calculate the luminosity distance (and corresponding comoving distance) at which this galaxy would have an observed magnitude of $m_r = 19.8$ mag, the GAMA limit.

(iii) Calculate the volume of the GAMA light-cone out to this comoving distance.

(iv) Divide by the total volume of our GAMA sample to get the selection probability.

(v) Multiply selection probabilities by 0.95 to account for the use of a 95 per cent complete sample based on K -corrections in GAMA (we do not attempt to model K -corrections for simulated galaxies).

(vi) Assign a random number to the halo and include the halo in our sample if this is less than the selection probability.

2.6 Comparison of group samples

Statistics for the groups selected in GAMA, the mocks and the comparison models are tabulated in Table 2. To complement this, the group mass distributions are shown in Fig. 4.

The GAMA group masses display a strongly peaked distribution, with more groups in $\mathcal{M}2$ and $\mathcal{M}3$ than the other bins. Comparing the halo and FoF mock groups, it is clear from the table that the FoF algorithm is systematically overestimating the numbers of groups for the two higher mass bins. The slightly lower mean mass of $\mathcal{M}1$ FoF versus halo groups is likely due to the fact that \mathcal{M}_{lum} is systematically underestimated for low redshifts where these low-mass haloes are found (see Fig. 1). For higher mass haloes, \mathcal{M}_{lum} correlates well with $\mathcal{M}_{\text{Dhalo}}$, (see Fig. 2), and so it seems likely that the higher numbers of larger mass FoF groups is due to the FoF algorithm aggregating lower mass haloes into one system.

Comparing the FoF mock groups with the GAMA groups, it is clear that the mock groups in the lowest mass bin tend to be of slightly lower mass than the corresponding GAMA groups, and of higher mass than the GAMA groups in the highest mass bin. It also appears that relatively there are slightly more high- than

low-mass groups in the FoF mocks. These differences should be borne in mind when comparing results from GAMA data and mock catalogues.

TNG matches the halo mock well on both the mean group masses and the mass distribution of selected haloes. The only apparent difference is in the relative numbers of groups in bin $\mathcal{M}4$ compared to the other bins, with TNG showing a greater relative number. This demonstrates the success of our group selection in TNG, which has the predominant effect of removing low-mass groups.

L-GALAXIES matches the halo mock mean group masses and follows very similar trends to TNG. It can be seen from Fig. 4 that the mass distribution is almost identical to that from TNG, except for a slightly greater number of haloes at the highest mass end.

3 MEASURING THE CORRELATION FUNCTION

We estimate the galaxy autocorrelation function and group–galaxy cross-correlation functions in bins of halo mass, as well as marked correlation functions, in which we weight groups and/or galaxies by their estimated mass.

We use CORRFUNC (Sinha & Garrison 2019, 2020) to calculate pair counts for the clustering statistics. When plotting correlation functions, we always plot w_p against the *mean separation* of galaxy pairs in each bin, rather than the centre of each (log-spaced) bin.

3.1 GAMA data and mock catalogues

In order to overcome the effects of redshift space distortions in the light-cones, we start by estimating the 2D group–galaxy cross-correlation function $\xi_{Gg}(r_{\perp}, r_{\parallel})$ and galaxy autocorrelation function $\xi_{gg}(r_{\perp}, r_{\parallel})$; the excess probability above random of finding a group and a galaxy (cross-correlation) or two galaxies (autocorrelation) separated by r_{\parallel} along the line of sight (LOS) and r_{\perp} perpendicular to the LOS. These separations are calculated using the standard method (e.g. Fisher et al. 1994) for pairs of objects with position vectors \mathbf{r}_1 and \mathbf{r}_2 . The separation is given by vector $\mathbf{s} = \mathbf{r}_2 - \mathbf{r}_1$ and the vector to the mid-point of the pair from an observer at the origin by $\mathbf{l} = (\mathbf{r}_1 + \mathbf{r}_2)/2$. The separations in the LOS and perpendicular directions are then given by $r_{\parallel} = |\mathbf{s} \cdot \hat{\mathbf{l}}|$, with $\hat{\mathbf{l}}$ being the unit vector in the direction of \mathbf{l} , and $r_{\perp} = \sqrt{|\mathbf{s}|^2 - r_{\parallel}^2}$.

Raw pair counts are obtained using CORRFUNC, then normalized to account for the relative total numbers of groups, N_G , galaxies, N_g , and random points, N_r . The normalized galaxy–galaxy, gg , group–galaxy, Gg , group–random, Gr , galaxy–random, gr , and random–random, rr , pair counts are then used to calculate the correlation functions. Specifically, these are obtained by dividing the raw pair counts in each separation bin by N_g^2 , $N_G N_g$, $N_G N_r$, $N_g N_r$, and N_r^2 , respectively.

The pair counts may additionally be weighted by group and/or galaxy mass in order to obtain marked correlation functions, and hence explore the dependence of clustering on group and galaxy mass. The random points, which follow the selection function of the galaxy sample, are generated as described in Section 2.4. A total of 426 790 random points are generated, 10 times the number of galaxies in the sample.

The galaxy autocorrelation $\xi_{gg}(r_{\perp}, r_{\parallel})$ is estimated using the standard Landy & Szalay (1993) estimator

$$\xi_{gg}(r_{\perp}, r_{\parallel}) = \frac{gg - 2gr + rr}{rr}, \quad (1)$$

while $\xi_{Gg}(r_\perp, r_\parallel)$ is estimated with the cross-correlation form (Mohammad et al. 2016) of this estimator

$$\xi_{Gg}(r_\perp, r_\parallel) = \frac{Gg - Gr - gr + rr}{rr}. \quad (2)$$

The 2D group–galaxy cross-correlation functions for our four mass bins of GAMA groups with our volume-limited sample of galaxies, are shown in Fig. 5. At small projected separations, $r_\perp \lesssim 5 h^{-1}$ Mpc, the clustering is seen to be stretched along the LOS direction (r_\parallel -axis). This is increasingly apparent in higher mass bins. At larger projected separations, the LOS clustering signal is compressed.

The projected auto- and cross-correlation functions, $w_p(r_\perp)$, are obtained by integrating the observed 2D correlation function $\xi(r_\perp, r_\parallel)$ along the LOS direction r_\parallel :

$$w_p(r_\perp) = 2 \int_0^{r_{\parallel\max}} \xi(r_\perp, r_\parallel) dr_\parallel. \quad (3)$$

We use a limit of $r_{\parallel\max} = 40 h^{-1}$ Mpc; following the results of Loveday et al. (2018, appendix B).

To estimate uncertainties on the clustering results from GAMA we use jackknife sampling. We use 27 regions in RA and calculate error bars as the square root of the diagonal terms in the covariance matrix calculated from these regions. For the mock catalogues nine different realizations are available and we estimate uncertainties using the scatter between these.

The jackknife sampling we use is designed to reproduce the cosmic variance between independent regions. This accurately reproduces the uncertainty on large scales, and on small scales can be interpreted as an upper bound on the variation between groups.

3.2 Simulations

TNG and L-GALAXIES use periodic boxes with no redshift space distortions, and so we can directly calculate the 3D correlation function $\xi(r)$ using the simplified formula

$$\xi(r) = \frac{DD}{RR} - 1, \quad (4)$$

with the normalized data pair count DD (Gg for the cross-correlation, gg for the autocorrelation) and random pair count RR. We again make use of CORRFUNC to calculate the data pair counts, normalized by total galaxy and group numbers as above.

Due to periodic boundary conditions, no random catalogue is needed. Instead, the normalized random pair count is calculated as

$$RR = \frac{v(r)}{V}, \quad (5)$$

where V is the total box volume and $v(r) = \frac{4}{3}\pi((r+dr)^3 - r^3)$ is the volume of a spherical shell of radius r and thickness dr (Alonso 2012).

The real-space 3D correlation function $\xi(r)$ is then converted to a projected correlation function using

$$\begin{aligned} w_p(r_\perp) &= 2 \int_0^{y_{\max}} \xi((r_\perp^2 + y^2)^{1/2}) dy \\ &= 2 \int_{r_\perp}^{r_{\max}} \frac{r \xi(r)}{\sqrt{r^2 - r_\perp^2}} dr, \end{aligned} \quad (6)$$

to produce a quantity directly comparable to the GAMA measurements. We perform this integral over an interpolation of the $\xi(r)$ and we again use an upper integration limit of $r_{\max} = 40 h^{-1}$ Mpc. It is pointed out in van den Bosch et al. (2013) that this integral may be biased on large scales relative to clustering calculated from

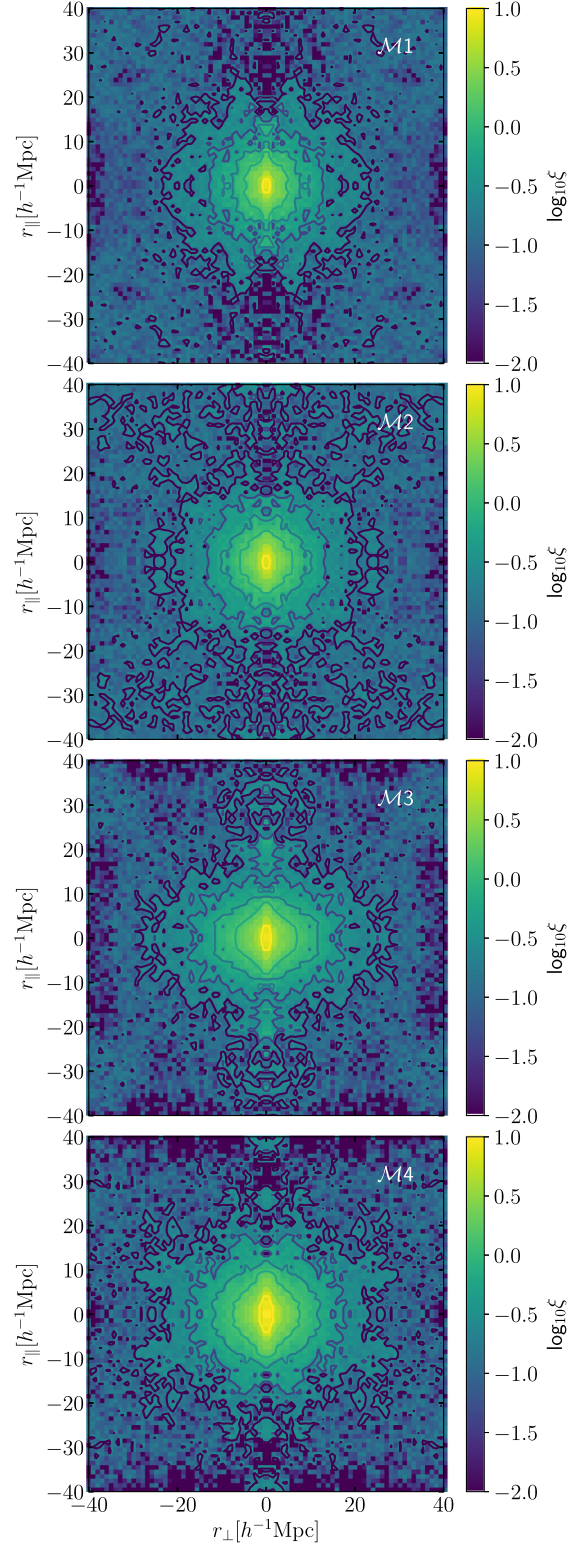


Figure 5. The 2D group–galaxy cross-correlation functions $\xi(r_\perp, r_\parallel)$ for our four bins of group mass. We show the clustering signal reflected about both axes to make it easier to see the distortions introduced by the peculiar velocities of galaxies around groups. Contour levels are the same as Li et al. (2006), going up from $\xi = 0.1875$ to $\xi = 48$ in factors of 2.

observations, but we do not attempt to correct for this as we are most interested in small scales.

To calculate uncertainties in the results for the simulation boxes we perform jackknife sampling by dividing the box into 27 subboxes and excluding these one at a time. We then give error bars as the square root of the diagonal elements of the covariance matrix. Jackknife sampling breaks the periodicity of the box, and should therefore require a random catalogue. However, we continue to use equation (5) for random pair counts, and account for the changed random–random term by scaling the $\xi(r)$ value in each jackknife region by the ratio of the overall $\xi(r)$ in the box against the mean $\xi(r)$ from the jackknife regions.

3.3 Marked correlation

The marked correlation M_w is calculated from the unweighted projected two-point correlation function w_p and weighted projected two-point correlation function W_p in all cases using (Sheth et al. 2005; Skibba et al. 2006)

$$M_w(r_\perp) = \frac{r_\perp + W_p(r_\perp)}{r_\perp + w_p(r_\perp)}. \quad (7)$$

Uncertainties on marked correlations would be overestimated if we simply combine the errors on W_p and w_p (see Skibba et al. 2006). Therefore, we calculate the marked correlation for each of our jackknife samples separately and estimate the uncertainty from these.

3.4 Bias

We make use of two bias measures in our analysis. The first is the relative bias of the group sample compared to the galaxy sample, which we define as

$$b_{\text{rel}}(r_\perp) = \frac{w_p^{\text{Gg}}(r_\perp)}{w_p^{\text{gg}}(r_\perp)}. \quad (8)$$

This accounts for different galaxy autocorrelation amplitudes between samples, although it does retain some dependence on the galaxy sample.

The second bias measure we use is that relative to dark matter. We define the galaxy bias b_g using

$$w_p^{\text{Gg}}(r_\perp) = b_g^2(r_\perp) w_p^{\text{DM}}(r_\perp), \quad (9)$$

and the corresponding group bias b_G with

$$w_p^{\text{Gg}}(r_\perp) = b_G(r_\perp) b_g(r_\perp) w_p^{\text{DM}}(r_\perp). \quad (10)$$

Note that in this notation the relative bias from equation (8) becomes $b_{\text{rel}} = b_G/b_g$.

For the dark matter autocorrelation, w_p^{DM} , we use the Millennium simulations, the Millennium (Springel et al. 2005) $\xi(r)$ on scales $r > 1 h^{-1}$ Mpc, and Millennium-II (Boylan-Kolchin et al. 2009) on smaller scales, which we project from $\xi(r)$ to $w_p(r_\perp)$ by interpolating $\xi(r)$ and using equation (6), the same method we used for the simulation correlations.

4 RESULTS

4.1 FoF versus halo mocks

We first compare clustering results obtained using the FoF and halo mocks in Fig. 6. We see that in mass bins 3 and 4, the FoF mock group

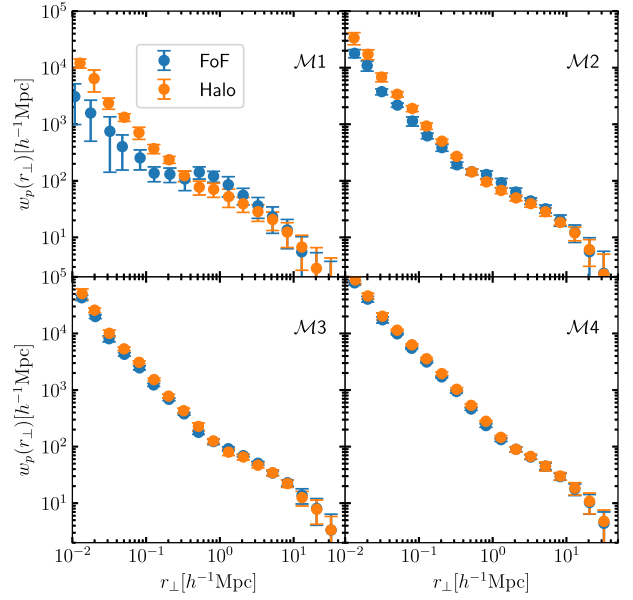


Figure 6. Group-galaxy cross-correlation functions for the mock catalogues. Orange symbols show results using the halo mocks, blue symbols show results obtained using FoF mocks.

clustering is in very good agreement with that of the halo mocks, despite the large excess of FoF groups in these mass bins (Table 2). However, for the lower mass bins, particularly $\mathcal{M}1$, the FoF group clustering is underestimated on very small scales, $r_\perp \lesssim 0.2 h^{-1}$ Mpc, and very slightly overestimated on scales $0.5 \lesssim r_\perp \lesssim 2 h^{-1}$ Mpc. It seems likely that the low-mass FoF groups may be contaminated by chance projections of isolated galaxies, thus reducing the small-scale clustering signal. Insofar as the mock catalogues are representative of the GAMA data, we can infer that the GAMA results are likely to be reliable in mass bins 2–4, but that those for $\mathcal{M}1$ should be treated with some scepticism.

To check the effects of the group finding on the marked correlation, we show the group mass marked correlation for the FoF and halo mocks in Fig. 7. We see that on small scales the mocks agree, but on scales $r_\perp \gtrsim 0.1 h^{-1}$ Mpc the FoF marked correlation is lower. This is around the size of a compact group, and is perhaps due both to spurious FoF groups (created by chance alignments) being isolated from other galaxies, and also to more extended groups being missed by the FoF group finder. We expect this trend to be representative of GAMA, and so the GAMA marked correlation may also be biased low on these scales.

4.2 Group clustering and bias in mass bins

4.2.1 GAMA and mocks

Fig. 8 shows the GAMA projected group–galaxy cross-correlation functions for each group mass bin (top), along with the bias relative to the galaxy sample (middle), and the bias relative to a DM-only simulation (bottom). Left-hand, middle, and right-hand panels show comparison results from the halo mocks, TNG, and L-GALAXIES, respectively. Both bias estimates are highly dependent on scale and group mass on intragroup scales, $r_\perp \lesssim 1 h^{-1}$ Mpc. On larger scales, the biases are relatively constant (within the error bars) for each mass bin, but there is still a slight trend for bias to increase with mass.

On scales $r_\perp \approx 0.1 h^{-1}$ Mpc, GAMA relative group bias (b_{rel} from equation 8; middle panels) increases rapidly with group mass, from b

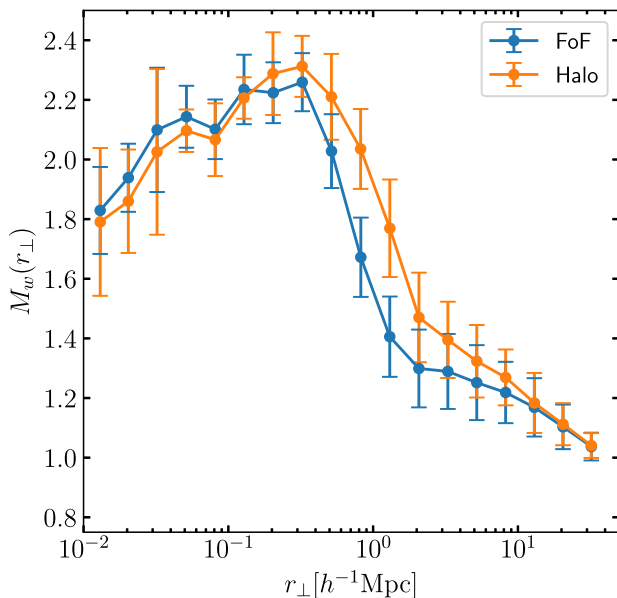


Figure 7. Marked correlation for the mock catalogues, using group mass as the mark. Orange symbols show results using the halo mocks, blue symbols show results obtained using FoF mocks.

$\approx 0.8 \pm 0.2$ for $\mathcal{M}1$ groups to $b \approx 5 \pm 1$ for $\mathcal{M}4$ groups. The strong halo-mass dependence of small-scale clustering seen here is to be expected, as on scales $r_{\perp} \lesssim 1 h^{-1}$ Mpc, the cross-correlation signal will be dominated by galaxies within each respective halo (intra-halo clustering) and group membership increases with halo mass.

Comparison in Fig. 6 of the FoF and halo mocks on these scales suggests that the apparent below-unity bias of $\mathcal{M}1$ groups is partly an artefact of the group-finding algorithm, although it also reflects a lack of bright galaxies in these small groups. The halo mock bias in the middle-left panel of Fig. 8 is consistent with unity for $\mathcal{M}1$ groups at $r_{\perp} \approx 0.1 h^{-1}$ Mpc, although it drops below unity above this, reaching a minimum at $r_{\perp} \approx 0.5 h^{-1}$ Mpc, indicating the spatial extent of these smaller groups. As with other mass bins, the mock galaxies in $\mathcal{M}1$ groups seem to be too centrally concentrated.

On larger scales ($1\text{--}5 h^{-1}$ Mpc), the dependence of relative bias on group mass is weaker, although the bias of the highest mass bin is still 2–3 times that of the lowest mass groups. By scales of $r_{\perp} \approx 10 h^{-1}$ Mpc, the biases of each mass bin are consistent within the uncertainties.

On the largest scales $r_{\perp} \gtrsim 10 h^{-1}$ Mpc, the relative bias remains constant in each bin within uncertainties but the GAMA auto- and cross-correlation functions are seen to have slightly greater amplitude than those of the mocks. This perhaps indicates small differences in the galaxy populations used, but as these scales are also the most affected by the projection of the clustering signal, we cannot draw any firm conclusions on these scales.

When turning to bias relative to the dark matter autocorrelation (b_g and b_G from equation (10); lower panels), the bias is seen to increase down to the smallest scales we plot for the galaxies and the groups in bins $\mathcal{M}2\text{--}4$. As with the bias relative to the galaxies, $\mathcal{M}1$ GAMA groups show a bias of about unity on the smallest scales not seen in the halo mock, which is likely to be a result of the group-finding algorithm.

The halo mocks substantially overpredict the bias on small scales. On intra-halo scales the relative bias (middle panels) is seen to increase roughly as a power law with decreasing r_{\perp} , rather than displaying a

flattening as seen in GAMA. This becomes even more apparent in bias relative to the dark matter (lower panels), with an even steeper increase to small scales. This suggests inaccuracy in the physics defining satellite galaxy occupation and positions in the mocks, with satellites being placed too close to the centre on average. This is perhaps unsurprising given the uncertainties in the modelling of satellite mergers when the dark matter subhalo they are associated with disappears (see e.g. Pujol et al. 2017).

4.2.2 TNG300 and L-GALAXIES

In Fig. 8, we also show corresponding results from the Illustris TNG300-1 simulation and the L-GALAXIES semi-analytic model, each around the mean GAMA redshift $z = 0.2$.

TNG results are shown as solid lines in the central column of panels. The TNG galaxy autocorrelation function (purple line) is in very close agreement with GAMA on scales $r_{\perp} \lesssim 5 h^{-1}$ Mpc, although slightly below that of the mock, and the TNG halo–galaxy cross-correlation functions show a similar characteristic inflection to GAMA around $r_{\perp} \approx 0.5\text{--}1 h^{-1}$ Mpc; the transition from the intra-halo to the inter-halo regime. In the higher mass bins, $\mathcal{M}3$ and $\mathcal{M}4$, the amplitude of the cross-correlations is also in agreement with GAMA on smaller scales within uncertainties. In $\mathcal{M}1$, and to a lesser extent in $\mathcal{M}2$, for which GAMA results are suspect, TNG shows a greater cross-correlation on scales $r_{\perp} \lesssim 0.3 h^{-1}$ Mpc than GAMA. This is clearest moving to the smallest scales, $r_{\perp} \lesssim 0.05 h^{-1}$ Mpc, where it leads to convergence of $\mathcal{M}1\text{--}\mathcal{M}3$ results as $\mathcal{M}1$ and $\mathcal{M}2$ continue to rise while $\mathcal{M}3$ and $\mathcal{M}4$ flatten off.

Solid lines in the right-hand panels of Fig. 8 show results for L-GALAXIES. Both the galaxy autocorrelation and halo–galaxy cross-correlations fall below the GAMA results. The relative biases in L-GALAXIES show the trend seen in the halo mock of a continuing increase down to the smallest scales and greater amplitude than GAMA, suggesting the same issues in the two SAMs. However, the group bias in the lower panels agrees well with GAMA on scales $r_{\perp} \gtrsim 0.1 h^{-1}$ Mpc, implying some of the discrepancy is connected to the galaxy sample. This difference in the dependence on the galaxy properties between L-GALAXIES and GAMA becomes clearer in the marked correlations discussed below. On larger scales, L-GALAXIES shows the halo mass dependence of bias continuing beyond $r_{\perp} = 5 h^{-1}$ Mpc, showing the most massive groups are at the centre of denser regions extending further than those of smaller groups, in agreement with GAMA.

4.3 Marked correlation functions

4.3.1 Marked cross-correlation

The upper panels of Fig. 9 show projected correlation functions weighted in the various ways indicated. Lower panels show marked group–galaxy cross-correlation functions using group mass (M_X^G), galaxy mass (M_X^g), and both masses (M_X^{Gg}) as weights, as well as the marked galaxy autocorrelation function (M_A^g). We weight by linear mass in order to enhance the differences between the marked statistics, although the use of log-mass weights does not qualitatively change our results (see Sheth et al. 2005 for a discussion on re-scaling marks). In Appendix D we show, using rank-ordered marks, that the specific values of the weights do not affect our conclusions.

The GAMA group-mass marked cross-correlation function (M_X^G , blue symbols) peaks at scales $r_{\perp} \approx 0.5 h^{-1}$ Mpc, declining gradually to smaller scales, and somewhat more rapidly on larger scales until $r_{\perp} \approx 2 h^{-1}$ Mpc, beyond which M_X^G declines more gradually.

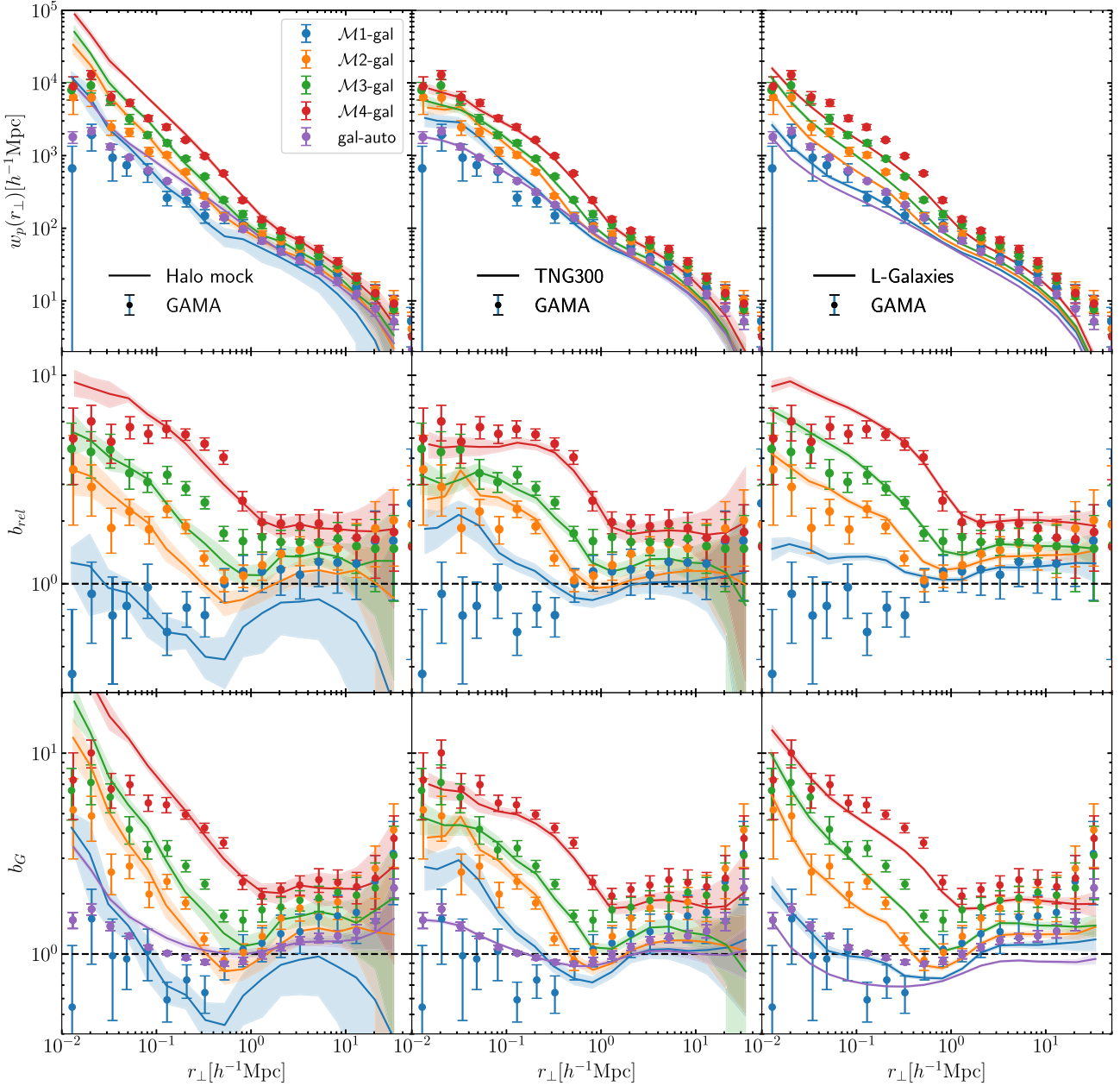


Figure 8. Top panels: The projected group–galaxy cross-correlation functions for our four bins of group mass as indicated. Also shown is the galaxy autocorrelation function. Middle panels: Relative bias of the projected group–galaxy cross-correlation to the galaxy sample, obtained by dividing the group–galaxy cross-correlation by the galaxy autocorrelation. Bottom panels: Bias of the projected group–galaxy cross-correlation and galaxy autocorrelation relative to the dark matter autocorrelation function of the Millennium simulations. In all panels, symbols and error bars show the GAMA results; lines of corresponding colour show results from the halo mock in the left-hand panels, the Illustris TNG300-1 simulation in the central panels, and L-GALAXIES in the right-hand panels.

The halo mock (blue line) shows similar trends to GAMA data, but with M_X^G about 20 per cent higher. The peak in M_X^G around $r_\perp \approx 0.5 h^{-1} \text{Mpc}$ is indicative of the typical projected radii of our galaxy groups. It is also consistent with the bias results of Fig. 8, where the relative strengths of the group biases differ most around this scale, due to the below-unity bias of $\mathcal{M}1$ groups and large bias of $\mathcal{M}4$ groups.

The GAMA galaxy-mass marked cross-correlation function (M_X^g , green points) is systematically greater than unity only on intergroup scales, $r_\perp \gtrsim 0.5 h^{-1} \text{Mpc}$. We are unable to measure M_X^g for the GAMA mocks, as galaxy masses are not available. When both galaxy and group masses are used as weights (M_X^{Gg} , orange points), a slight

additional enhancement is seen relative to M_X^G , indicative of the most massive groups having an enhanced number of massive satellite galaxies.

M_X^G measurements from both the TNG and L-GALAXIES simulations show general agreement with GAMA. TNG agrees with GAMA within uncertainties on almost all scales, but is below the mocks on scales $r_\perp \lesssim 1 h^{-1} \text{Mpc}$. L-GALAXIES on the other hand agrees well on all scales with the halo mock, and is generally above but just consistent with the GAMA results. The very close agreement between L-GALAXIES and the halo mock may be a result of both being built upon the Millennium simulation.

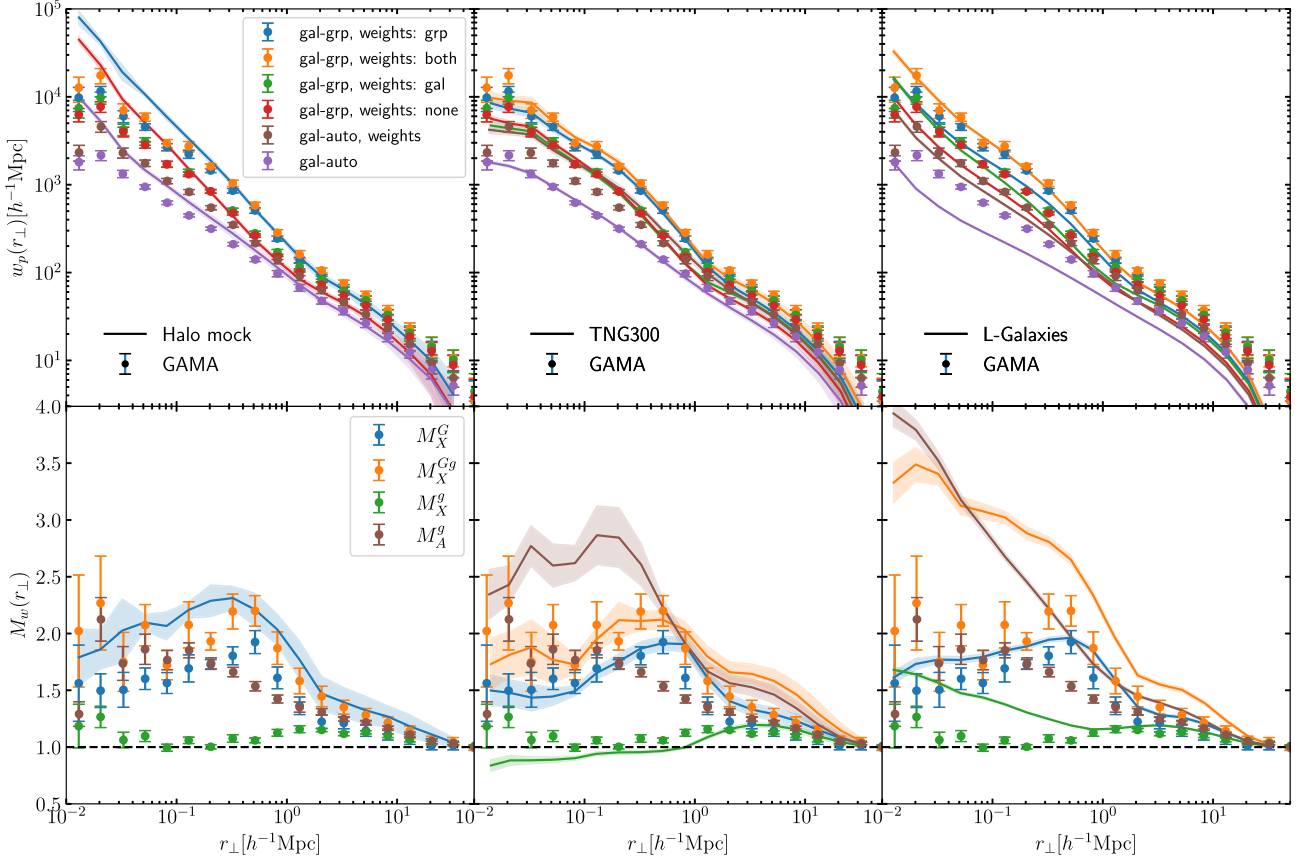


Figure 9. Top panels: The projected group–galaxy cross-correlation functions for all groups, weighted by galaxy and group masses as indicated. Also shown is the galaxy autocorrelation function both unweighted and using galaxy masses as weights. Bottom panels: Marked cross-correlations using galaxy masses (M_X^g), group masses (M_X^G), and both masses (M_X^{Gg}) as marks, along with the stellar-mass marked galaxy autocorrelation (M_A^g). In all panels, symbols and error bars show the GAMA results; lines of corresponding colour show results from the halo mock in the left-hand panels, the Illustris TNG300-1 simulation in the central panels and L-GALAXIES in the right-hand panels.

When marking with galaxy masses, TNG shows $M_X^g < 1$ on scales $r_\perp \lesssim 0.5 h^{-1} \text{ Mpc}$, meaning the most massive satellite galaxies are not found near the group centres. Yet when both group and galaxy masses are used (M_X^{Gg}), an enhancement relative to M_X^G is seen on all scales. This is consistent with the conclusion from GAMA that the most massive groups also contain the most massive satellites, but this dependency extends out slightly further in TNG, to $r_\perp \approx 10 h^{-1} \text{ Mpc}$.

L-GALAXIES shows a galaxy-mass marked cross-correlation M_X^g greater than unity, especially on scales $r_\perp \lesssim 1 h^{-1} \text{ Mpc}$ where M_X^g is seen to increase as scale decreases, meaning massive satellites are always closely associated with the group centre. The same trend is seen and enhanced even further when both group and galaxy masses are used as marks (M_X^{Gg}). This is consistent with the high small-scale bias we observed for L-GALAXIES, yet very different from the GAMA result, suggesting that the satellite galaxies in L-GALAXIES are typically more massive. This is in accord with the finding in VM20 that the modified Schechter functions appropriate for GAMA satellite galaxies underpredict the number of massive satellites in L-GALAXIES.

4.3.2 Marked autocorrelation

For GAMA, L-GALAXIES, and TNG, we also show the (stellar mass) marked galaxy autocorrelation (M_A^g , brown symbols or lines), which helps in understanding some of the differences in the group-galaxy

cross-correlations. GAMA shows no systematic scale-dependence (but large scatter) in M_A^g on scales $r_\perp \lesssim 0.2 h^{-1} \text{ Mpc}$, but then declines systematically on larger scales, always lying below M_X^{Gg} . This makes sense, as M_A^g indirectly contains group information through the presence of central galaxies, although these will have lower masses than the groups.

TNG on the other hand shows a marked autocorrelation M_A^g which peaks on scales $0.1\text{--}0.5 h^{-1} \text{ Mpc}$ and decreases slightly on smaller scales. The large enhancement compared to GAMA and the TNG cross-correlation functions is likely to be due to the apparent overdependence of central galaxy mass on group mass in TNG reported by VM20. The decreasing dependence on the smallest scales is consistent with the trends in M_X^g , and shows that the most massive galaxies have a slight tendency to avoid group centres.

L-GALAXIES shows a very different trend that the most massive galaxies are very close together, with M_A^g still increasing at $r_\perp \approx 0.01 h^{-1} \text{ Mpc}$. This matches the cross-correlation result and also appears consistent with a slight trend in Henriques et al. (2017) for the autocorrelation to be below SDSS in lower mass bins and above in higher mass bins. This is likely to be the result of the supernova feedback used, as van Daalen et al. (2016) find that the feedback strength affects the relative proportions of satellite galaxies of different masses.

The general picture found from the marked correlations is one of agreement in the group mass dependence of clustering, but

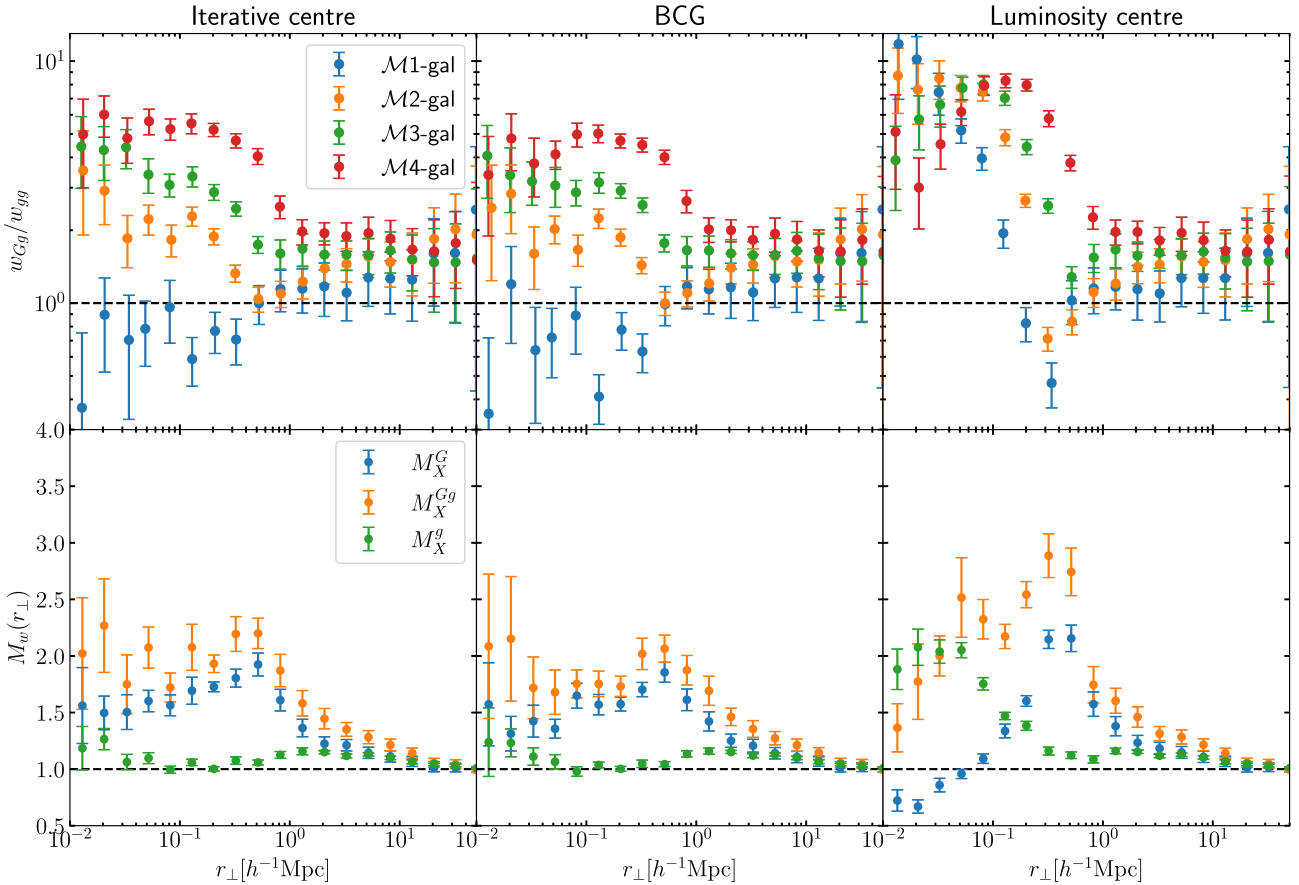


Figure 10. Effect of choice of GAMA group centre on our results. Upper panels show the relative bias b_{rel} of the projected group–galaxy cross-correlation to the galaxy sample, and lower panels show the marked correlation using galaxy masses (M_X^g), group masses (M_X^G), and both masses (M_X^{Gg}) as marks. The left-hand panels show the iterative group centre, the middle panels the brightest central galaxy, and the right-hand panels the centre of light of the group.

disagreement in the galaxy mass dependence. While the group mass dependence is a significant success in the positioning of galaxies within groups in both TNG and L-GALAXIES, massive galaxies appear to be too clustered, especially in L-GALAXIES.

5 DISCUSSION

To put our results into context we discuss here the choice of group centre, which is the main caveat to our work, and compare against previous works.

5.1 Choice of group centre

In this work, we have considered group–galaxy cross-correlation functions in GAMA down to scales smaller than the typical group size, so our results depend heavily on the choice of group centre. We check here for effects due to possible mis-identification of group centre by using the three different definitions of group centre described in section 4.2.1 of R11.

R11 found the most reliable group centre to be the one we have used throughout this work, the iterative centre. This was found by iteratively removing the galaxy furthest from the centre of light of all remaining galaxies in the group, until only one galaxy remains. The position of the final galaxy is taken to be the group centre. In most cases this is the same as the second definition of group centre, the brightest central galaxy (BCG), taken to be the brightest galaxy

in the group. The third definition of group centre corresponds simply to the group centre of light, which does not in general coincide with a galaxy. Using mock catalogues, R11 showed the iterative centre to match the true centre in ~ 90 per cent of cases, while the BCG showed large offsets in some cases, and the centre of light only matched the true centre for groups where all members are detected.

To explore the effect of group centre choice on our results, we show in Fig. 10 the relative bias b_{rel} of the four group mass bins and the marked cross-correlations for the three definitions of group centre. On the left we show the iterative centre used elsewhere in this work. This is in most cases the same as the BCG shown in the middle panel, so the results are similar from these two options. However, the iterative centre shows a more consistent picture for different group masses on small scales, while the BCG shows a drop in bias for the most massive groups, suggesting the galaxy at the centre of the gravitational potential of the group has been included in the cross-correlation. The definition of group centre as centre of light is shown in the right-hand panel, and this definition shows significant evidence of miscentring. The bias is seen to be peaked, with the peak at $r_{\perp} \approx 0.1 h^{-1} \text{ Mpc}$ for the most massive groups, and on smaller scales for less massive groups. The location of this peak is indicative of the mean offset of the central galaxy from the centre of light.

A similar outcome is found by considering the marked correlations. Using group mass as the mark, the iterative centre and BCG results are similar, but the centre of light definition shows a negative mark on small scales related to the reduction in bias for the more

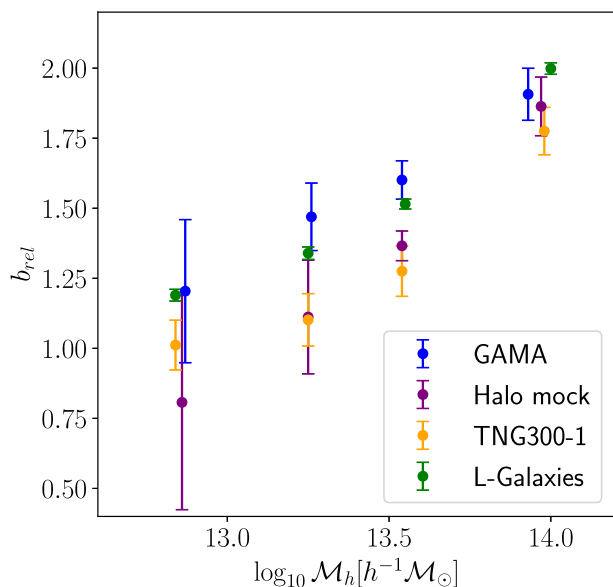


Figure 11. Large-scale relative bias, averaged over scales 2–10 h^{-1} Mpc, as a function of halo mass for GAMA, the halo mocks, TNG300-1 and L-GALAXIES. Uncertainties are calculated by jackknife of the average b_{rel} for GAMA, TNG and L-GALAXIES, and by scatter between the 9 realizations for the halo mocks.

massive groups. When using galaxy masses as marks, the iterative centre and BCG results both show no mark on scales less than the typical group size, but the centre of light shows a positive mark, probably indicating the inclusion of the true central galaxy in the cross-correlation.

Based on this, we are in agreement with the result of R11 that the iterative centre we have used is the best reflection of the true group centre, as it does not display the offset in peak bias associated with including the central galaxy in the cross-correlation.

5.2 Comparison with previous results

Finally, we compare our results to previous works, and calculate the average bias on large scales.

Our finding of an increase in clustering amplitude with group mass on scales of a few h^{-1} Mpc agrees with the results from the analysis of SDSS data by Wang et al. (2008). These authors found that the bias relative to the lowest mass bin increases quadratically with mass, and we show a similar rise in our relative bias in Fig. 11, with bias averaged over scales 2–10 h^{-1} Mpc. This trend is consistent with the results from the simulations, albeit with a slightly higher normalization. However, due to our use of different, narrower, mass bins than Wang et al. (2008), the uncertainties from GAMA are large, and the bias values are not directly comparable. In addition to the large-scale bias, we show on smaller, intragroup, scales, which were not considered by Wang et al. (2008), that the dependence of clustering amplitude on group mass become significantly stronger.

This sharp increase in cross-correlation amplitude within the typical group radius matches the results of Berlind et al. (2006), as does evidence for a flattening of the cross-correlation on scales $r_{\perp} \lesssim 0.3 h^{-1}$ Mpc in our GAMA and TNG results. Berlind et al. (2006) attribute this to either a core to the radial profile of satellite galaxies, or to misidentification of the centre. We do not find evidence that the central galaxies are incorrect in our data, so support the explanation of a central core to groups.

The result from the marked correlation functions that massive galaxies are associated with massive groups is not surprising, and consistent with GAMA results from VM20. More interesting is the lack of dependence of the mark on galaxy mass alone within the radii of the smallest groups, in agreement with the results of Kafle et al. (2016) that there is no mass segregation within GAMA groups. This is in contrast to the results from SDSS, most recently in Roberts et al. (2015), that more massive satellites are generally closer to the group centre. Our approach of using the marked correlation is a new method to test for mass segregation, but as our galaxy sample is volume limited in r -band luminosity and not in mass, our results are not directly comparable to these previous studies, and the marked correlations must be interpreted with caution given that stellar masses increase with group mass.

The lack of mass segregation also suggests a breakdown in self-similarity on group scales, as the most massive groups are found to be the most clustered, but this trend does not continue to galaxies within the groups. This suggests that while on intergroup scales the galaxy distribution depends primarily on the dark matter distribution, within groups, baryon astrophysics has a significant effect.

6 CONCLUSIONS

In this work we present group–galaxy cross-correlation functions and mass-weighted marked correlations for the GAMA survey, GAMA mocks, the TNG300-1 simulation, and the L-GALAXIES semi-analytic model. We use four group mass bins with $12.0 < \lg \mathcal{M}_h < 14.8$ and cross-correlate with a volume-limited galaxy sample with density $5.38 \times 10^{-3} h^3 \text{ Mpc}^{-3}$.

We find that the group–galaxy cross-correlation function (Fig. 8) increases systematically with group mass and with decreasing scale below $r_{\perp} \approx 1 h^{-1}$ Mpc. There is no scale dependence on scales $r_{\perp} \gtrsim 1 h^{-1}$ Mpc, but the correlation amplitude still increases with group mass, indicating that more massive groups are embedded within extended overdense structures.

Using marked correlations (Fig. 9), we see that the cross-correlation has the strongest group mass dependence at scales $r_{\perp} \approx 0.5 h^{-1}$ Mpc, the typical group radius (defined as projected separation to the most distant member galaxy from the group centre). No direct dependence on galaxy mass is observed, but the combination of group and galaxy mass causes an enhancement over the use of group mass only. This leads us to conclude that massive satellite galaxies are generally found in massive groups, but do not preferentially lie close to the central galaxy. Note that the central galaxy coincides with the iterative group centre, and so central-group pairs are not included in the group–galaxy cross-correlation functions presented.

6.1 Comparison to mocks and simulations

We use the GAMA mock catalogues to explore the effects of systematics in the data, particularly the group mass estimates, and to examine the model used for the mocks. Comparison of mocks using FoF and halo based group finding methods suggests that the masses may be overestimated at high redshift and underestimated at low redshift, although this only causes differences in the cross-correlation function in our lowest mass bin, $\mathcal{M}1$.

We have also compared our results against the TNG300-1 box from the IllustrisTNG hydrodynamical simulation and to the L-GALAXIES semi-analytic model. In order to provide a fair comparison, we selected groups using a simple model of the GAMA selection function.

The IllustrisTNG hydrodynamical simulation agrees well with our GAMA results in all cross-correlation bins except the lowest mass bin where the GAMA results are least reliable. It also displays very similar marked cross-correlations to GAMA, evidencing accuracy in the distribution of galaxies around groups. The only significant difference between TNG and GAMA we see is in the marked galaxy autocorrelation, where the enhancement in TNG appears to be the same overdependence of central galaxy mass on group mass seen in VM20.

The L-GALAXIES model is found to overpredict the mass dependence of the cross-correlation, showing an increasing bias down to the smallest scales considered. This is seen in the marked correlations to be driven by stronger clustering than GAMA of the most massive galaxies, perhaps driven by inaccurate supernova feedback. Together with the difficulties of modelling the infall of satellites without surviving subhaloes, this results in too many galaxies in the inner parts of the haloes. Away from the group centre, L-GALAXIES shows similar group bias to GAMA, demonstrating that the distribution of galaxies in the outer regions of the haloes is realistic.

6.2 Future prospects

While the GAMA groups are expected to be more reliable than the SDSS groups used in previous works, due to high spectroscopic completeness and the use of only the most reliable groups with $N_{\text{FoF}} \geq 5$, we are limited by the smaller area of the GAMA survey. In future, the Wide Area VISTA Extragalactic Survey (Driver et al. 2019) is expected to be able to produce a much larger sample of galaxy groups and so improve upon our results by reducing the uncertainties and allowing the use of finer mass bins.

ACKNOWLEDGEMENTS

SDR is supported by a Science and Technology Facilities Council (STFC) studentship. RYMB was supported by a STFC studentship. JL acknowledges support from the STFC (grant number ST/I000976/1).

We thank Michael Boylan-Kolchin for providing the Millennium and Millennium-II simulation dark matter correlation functions in machine-readable form.

GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programs including GALEX MIS, VST KiDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT, and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is <http://www.gama-survey.org/>.

We acknowledge the IllustrisTNG team for making their simulation data available. This work used the 2015 public version of the Munich model of galaxy formation and evolution: L-Galaxies. The source code and a full description of the model are available at <https://lgalaxiespublicrelease.github.io/>.

Finally, we thank the anonymous referee for providing helpful comments about the manuscript.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. Tabulated clustering results will be made available via the GAMA website <http://www.gama-survey.org/>.

REFERENCES

- Adelman-McCarthy J. K. et al., 2006, *ApJS*, 162, 38
 Alonso D., 2012, preprint ([arXiv:1210.1833](https://arxiv.org/abs/1210.1833))
 Armijo J., Cai Y.-C., Padilla N., Li B., Peacock J. A., 2018, *MNRAS*, 478, 3627
 Beisbart C., Kerscher M., 2000, *ApJ*, 545, 6
 Berlind A. A., Kazin E., Blanton M. R., Pueblas S., Scoccimarro R., Hogg D. W., 2006, preprint ([arXiv:astro-ph/0610524](https://arxiv.org/abs/astro-ph/0610524))
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150
 Budzynski J. M., Koposov S. E., McCarthy I. G., McGee S. L., Belokurov V., 2012, *MNRAS*, 423, 104
 Driver S. P. et al., 2009, *Astron. Geophys.*, 50, 5.12
 Driver S. P. et al., 2011, *MNRAS*, 413, 971
 Driver S. P. et al., 2019, *The Messenger*, 175, 46
 Eke V. R., Baugh C. M., Cole S., Frenk C. S., Navarro J. F., 2006, *MNRAS*, 370, 1147
 Farrow D. J. et al., 2015, *MNRAS*, 454, 2120
 Fisher K. B., Davis M., Strauss M. A., Yahil A., Huchra J., 1994, *MNRAS*, 266, 50
 Girardi M., Giuricin G., Mardirossian F., Mezzetti M., Boschin W., 1998, *ApJ*, 505, 74
 Gunawardhana M. L. P. et al., 2018, *MNRAS*, 479, 1433
 Hamilton A. J. S., Tegmark M., 2004, *MNRAS*, 349, 115
 Han J. et al., 2015, *MNRAS*, 446, 1356
 Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, *MNRAS*, 367, 1039
 Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *MNRAS*, 451, 2663
 Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Wang W., 2017, *MNRAS*, 469, 2626
 Hopkins A. M. et al., 2013, *MNRAS*, 430, 2047
 Humason M. L., Mayall N. U., Sandage A. R., 1956, *AJ*, 61, 97
 Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, *MNRAS*, 440, 2115
 Kaffle P. R. et al., 2016, *MNRAS*, 463, 4194
 Knobel C. et al., 2012, *ApJ*, 755, 48
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
 Li C., Jing Y. P., Kauffmann G., Börner G., White S. D. M., Cheng F. Z., 2006, *MNRAS*, 368, 37
 Lilly S. J. et al., 2007, *ApJS*, 172, 70
 Liske J. et al., 2015, *MNRAS*, 452, 2087
 Loveday J. et al., 2015, *MNRAS*, 451, 1540
 Loveday J. et al., 2018, *MNRAS*, 474, 3435 (L18)
 Marinacci F. et al., 2018, *MNRAS*, 480, 5113
 Mo H. J., White S. D. M., 1996, *MNRAS*, 282, 347
 Mohammad F. G., de la Torre S., Bianchi D., Guzzo L., Peacock J. A., 2016, *MNRAS*, 458, 1948
 Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
 Nelson D. et al., 2018, *MNRAS*, 475, 624
 Nelson D. et al., 2019, *Comput. Astrophys. Cosmol.*, 6, 2
 Pillepich A. et al., 2018, *MNRAS*, 475, 648
 Presotto V. et al., 2012, *A&A*, 539, A55
 Press W. H., Schechter P., 1974, *ApJ*, 187, 425
 Pujol A. et al., 2017, *MNRAS*, 469, 749
 Roberts I. D., Parker L. C., Joshi G. D., Evans F. A., 2015, *MNRAS*, 448, L1
 Robotham A. S. G. et al., 2011, *MNRAS*, 416, 2640 (R11)
 Seljak U., Warren M. S., 2004, *MNRAS*, 355, 129
 Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
 Sheth R. K., Tormen G., 2004, *MNRAS*, 350, 1385
 Sheth R. K., Mo H. J., Tormen G., 2001, *MNRAS*, 323, 1
 Sheth R. K., Connolly A. J., Skibba R., 2005, *MNRAS*, preprint ([arXiv:astro-ph/0511773](https://arxiv.org/abs/astro-ph/0511773))
 Sinha M., Garrison L. H., 2019, Software Challenges to Exascale Computing. Second Workshop. p. 3
 Sinha M., Garrison L. H., 2020, *MNRAS*, 491, 3022
 Skibba R., Sheth R. K., Connolly A. J., Scranton R., 2006, *MNRAS*, 369, 68

- Skibba R. A., Sheth R. K., Croton D. J., Muldrew S. I., Abbas U., Pearce F. R., Shattow G. M., 2013, *MNRAS*, 429, 458
- Springel V. et al., 2005, *Nature*, 435, 629
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Stoyan D., Stoyan H., 1994, *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. Wiley, Chichester
- Sureshkumar U. et al., 2021, *A&A*, preprint (arXiv:2102.04177)
- Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, *MNRAS*, 387, 1391
- Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, *ApJ*, 631, 41
- van Daalen M. P., Henriques B. M. B., Angulo R. E., White S. D. M., 2016, *MNRAS*, 458, 934
- van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, *MNRAS*, 430, 725
- Vázquez-Mata J. A. et al., 2020, *MNRAS*, 499, 631 (VM20)
- Viola M. et al., 2015, *MNRAS*, 452, 3529
- von der Linden A., Wild V., Kauffmann G., White S. D. M., Weinmann S., 2010, *MNRAS*, 404, 1231
- Wang Y., Yang X., Mo H. J., van den Bosch F. C., Weinmann S. M., Chu Y., 2008, *ApJ*, 687, 919
- Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, *MNRAS*, 432, 336
- White M., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 057
- White M., Padmanabhan N., 2009, *MNRAS*, 395, 2381
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, *ApJ*, 671, 153

APPENDIX A: GALAXY SAMPLE STATISTICS

We desire our volume-limited GAMA, mock, TNG, and L-GALAXIES galaxy samples to have comparable clustering statistics. In order to achieve this, they were defined to have similar number densities (Table 1). Here, we show the stellar mass distributions and autocorrelation functions of these samples.

The distributions of stellar masses in each sample (Fig. A1) show some variation. This is not surprising, as the samples are volume-limited in r -band luminosity and not in mass, and so variations in mass-to-light ratio will affect mass completeness. Compared to the GAMA sample, TNG shows a narrower peak but an overabundance

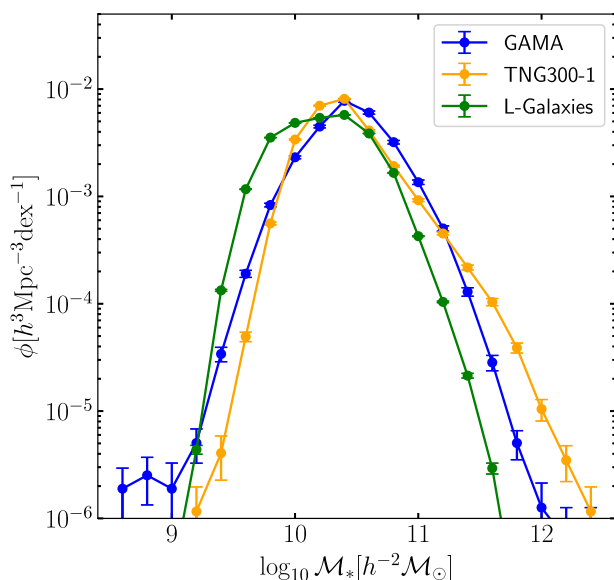


Figure A1. Distribution of stellar masses in our galaxy samples from GAMA, TNG, and L-GALAXIES. The mock catalogues do not include stellar masses so are not shown.

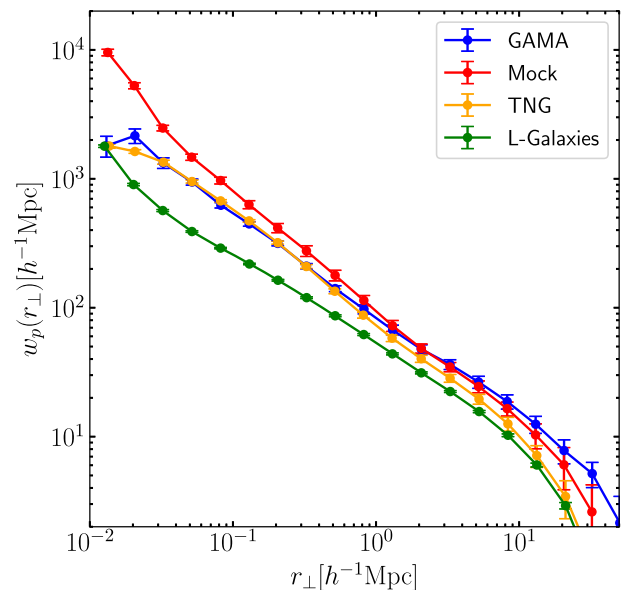


Figure A2. Projected autocorrelation functions of our galaxy samples from GAMA, the mock catalogues, TNG, and L-GALAXIES.

of the most massive galaxies with $\log_{10} \mathcal{M}_* \gtrsim 11.2 h^{-2} \mathcal{M}_{\odot}$. L-GALAXIES shows a shift to slightly smaller masses than GAMA.

Fig. A2 shows the projected autocorrelation functions of the galaxy samples. On small scales, GAMA and TNG agree well but the mocks show a slightly greater autocorrelation and L-GALAXIES shows a lower autocorrelation. On the largest scales GAMA shows the greatest clustering, but consistent within uncertainties with the mocks.

APPENDIX B: GROUP SELECTION IN SIMULATIONS

Here, we compare four methods of selecting groups in mass bins from the TNG and L-GALAXIES simulations, and the effect these methods have on estimated relative bias. The four group selection methods compared are:

- (i) Random sampling to mimic GAMA group selection, the method described in Section 2.5 and used elsewhere.
- (ii) Spatial sampling to mimic GAMA group selection. Here, we select groups within a distance from the origin corresponding to the comoving distance at which the fifth brightest member galaxy would have an apparent magnitude of $m_r = 19.8$. This removes the periodicity of the box, and we therefore calculate the correlation function using the full Landy & Szalay (1993) estimator with random galaxies distributed around the box. Uncertainties on this sample are estimated using jackknife between 27 samples of equal volume selected by angle, and are larger than those of the random selection due to the loss of periodicity.
- (iii) Use only of GAMA mass bin limits, without further selection. This results in an overabundance of low-mass groups in a volume-limited simulation cube compared to GAMA.
- (iv) Adjustment of mass limits to match the mean group masses in GAMA (the method employed in VM20).

Comparing these different selection methods applied to TNG in Fig. B1, the relative bias is consistent between the samples selected using methods (i) and (ii), except for the smallest scales in \mathcal{M}_1 . Bearing in mind that the groups in sample (i) are randomly distributed throughout the TNG data cube, whereas those in sample (ii) lie

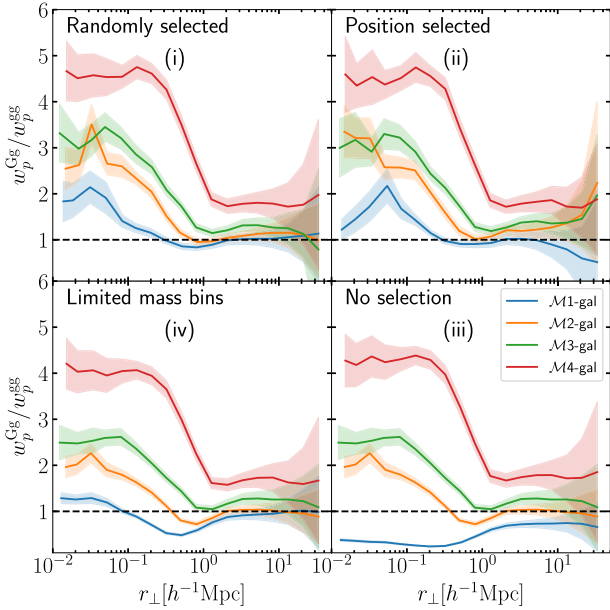


Figure B1. Relative bias for the 4 mass bins in the TNG simulations using different selection options for groups. Clockwise from top left, the panels show the four group selections (i)–(iv): upper left, the selection of groups throughout the volume based on galaxy luminosities used in this work; upper right, a group selection based on galaxy luminosities and radial distance from box origin; lower left, the full group sample with low and high mass groups removed to match GAMA mean masses; and lower right, the full group sample in the volume-limited simulation.

predominantly closer to the origin, this comparison illustrates that the spatial selection of the groups has only minimal effect on the group–galaxy cross-correlation function, and justifies our choice of random sampling (method i). The differences in very small-scale clustering in $\mathcal{M}1$ likely arise from sampling fluctuations, since the sample (ii) TNG $\mathcal{M}1$ groups are only taken from approximately 10 per cent of the total volume.

Sample (iii), lower right panel, shows very different results. The addition of many low-mass groups forces the bias for the lower mass bins down, leading to antibias on all scales for $\mathcal{M}1$ and near the group edge for $\mathcal{M}2$. This is likely due to the $\mathcal{M}1$ TNG central galaxies having a mean luminosity ≈ 0.2 mag lower than the comparison galaxy sample. Using sample (iv), lower left panel, increases the bias for $\mathcal{M}1$ but it still remains below that of sample (i).

The comparison of these selection methods has demonstrated the importance of mimicking the selection function in GAMA and validated our approach to doing so.

APPENDIX C: EFFECT OF GROUP SELECTION ON THE CROSS-CORRELATION

We show here that our GAMA cross-correlation results are not significantly affected by group selection effects. Fig. C1 shows the cross-correlation for the $\mathcal{M}4$ bin in the FoF mock with different artificial selection effects introduced.

To check the effects of missing groups near the field edges, we select groups based on the distance from the field centres. This results in a slight increase in cross-correlation amplitude on large scales, but consistent within uncertainties. We also show the effects of selecting low- and high-redshift groups. There are no significant shifts in either case.

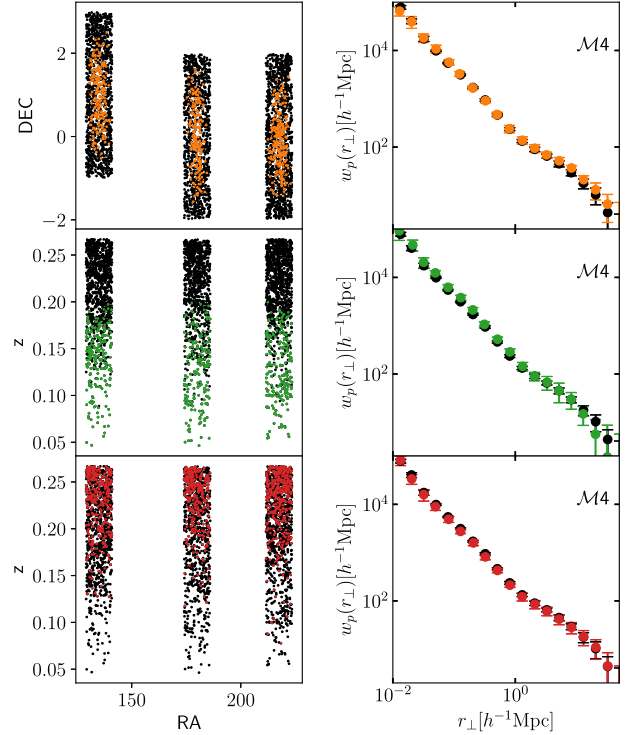


Figure C1. The effect of group selection on the group–galaxy cross-correlation function in the FoF mock catalogue for bin $\mathcal{M}4$. Left-hand panels show the selected groups and right-hand panels show the resulting cross-correlation, with black points in all cases showing the full sample. The upper row shows a selection excluding groups near the field edges, the middle panels show low-redshift groups, and the lower panels high-redshift groups.

The similarity of all the cross-correlations shown here (and similar results are obtained for the other mass bins and the halo mock) demonstrates that our results are robust to the effects of group selection.

APPENDIX D: MARKED CORRELATIONS BY RANK

In order to check the effect of our choice of galaxy or group mass as a mark, we perform an alternative marking using the rank ordering method of Skibba et al. (2013).

We sort the masses in ascending order and assign the rank as the position in the sorted list. Results from using these ranks as marks are shown in Fig. D1. When compared to the marked correlations using masses shown in Fig. 9, it is clear that the amplitude of the marked correlations is reduced when using ranks. However, the qualitative comparison between different weighting options and samples remains the same.

The most notable difference is the TNG galaxy mass-weighted autocorrelation. In that case, using rank orderings brings the mark into agreement with GAMA on most scales, suggesting that the enhanced mark seen in Fig. 9 is due to the differences in the shape of the stellar mass function between TNG and GAMA in Fig. A1.

The other visible difference is that the cross-correlation weighted by galaxy masses is greater than 1 when using ranks for GAMA and TNG. However, there is no scale dependence, meaning this is not a signal of mass segregation. Instead, it appears to confirm the galaxies from our sample which are in the groups have slightly higher masses than the average of the volume-limited sample.

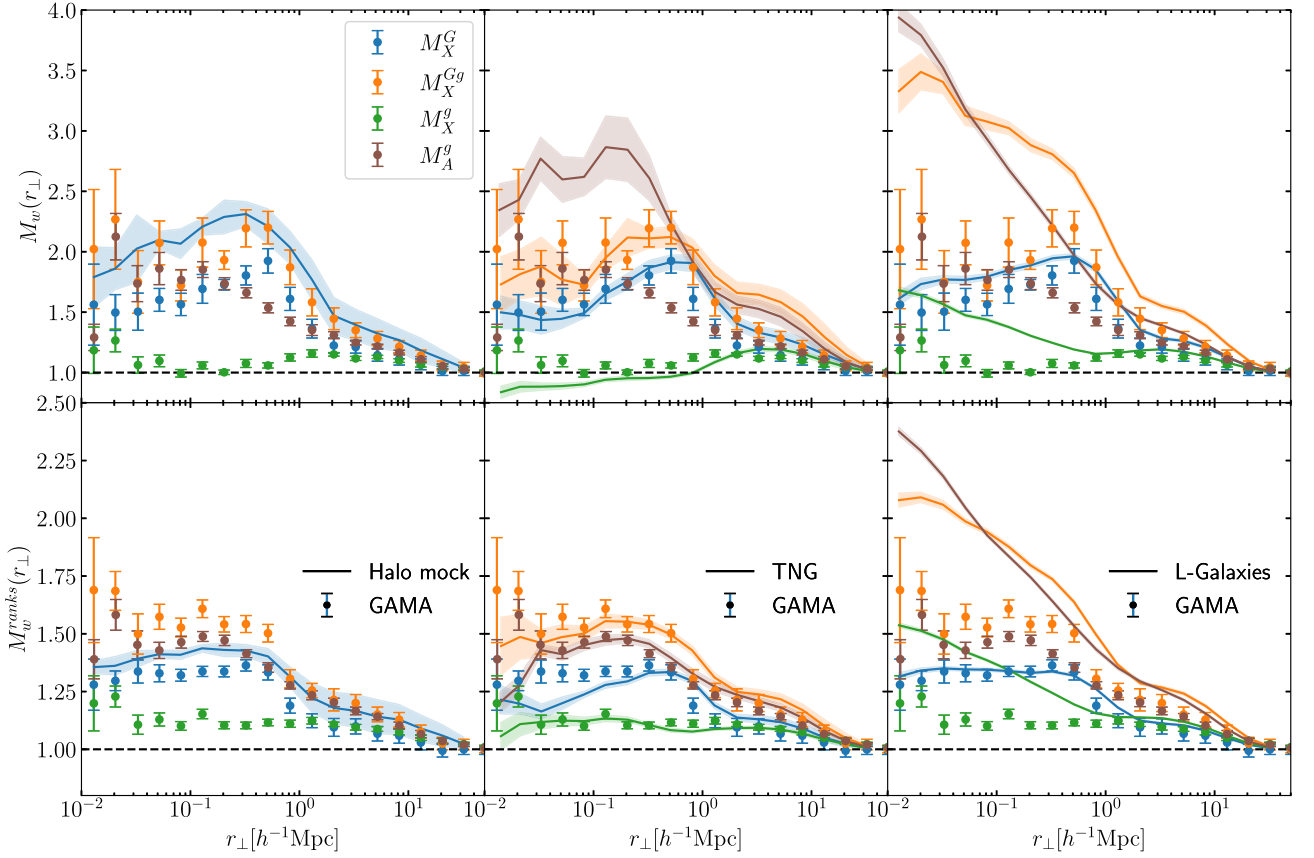


Figure D1. Marked correlations using masses and the rank ordering of mass. Upper panels are the same as the lower panels of Fig. 9, lower panels show the results from rank ordering the masses. Symbols and error bars show the GAMA results in all panels; lines of corresponding colour show results from the halo mock in the left-hand panels, the Illustris TNG300-1 simulation in the central panels and L-GALAXIES in the right-hand panels.

This paper has been typeset from a \LaTeX file prepared by the author.